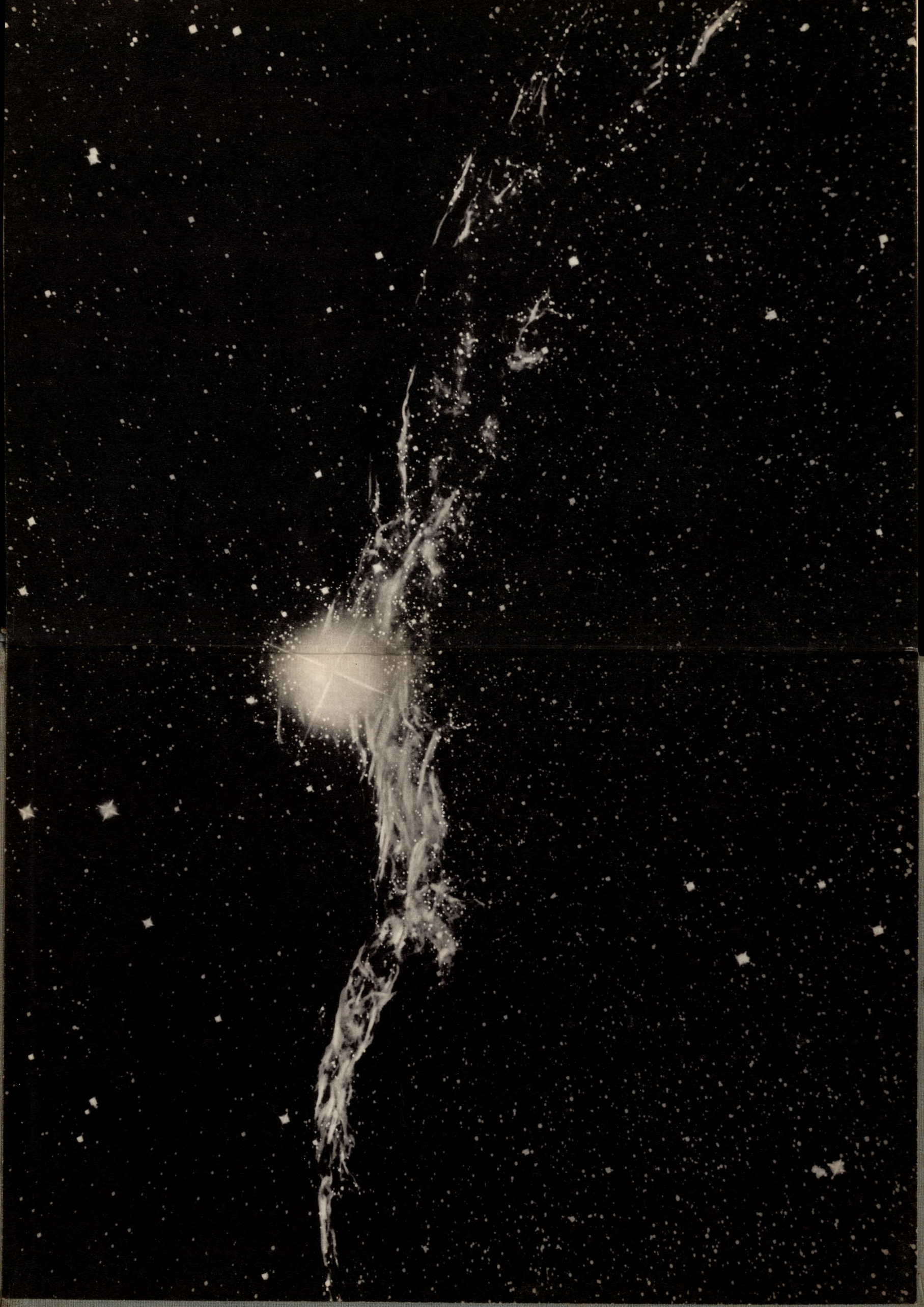
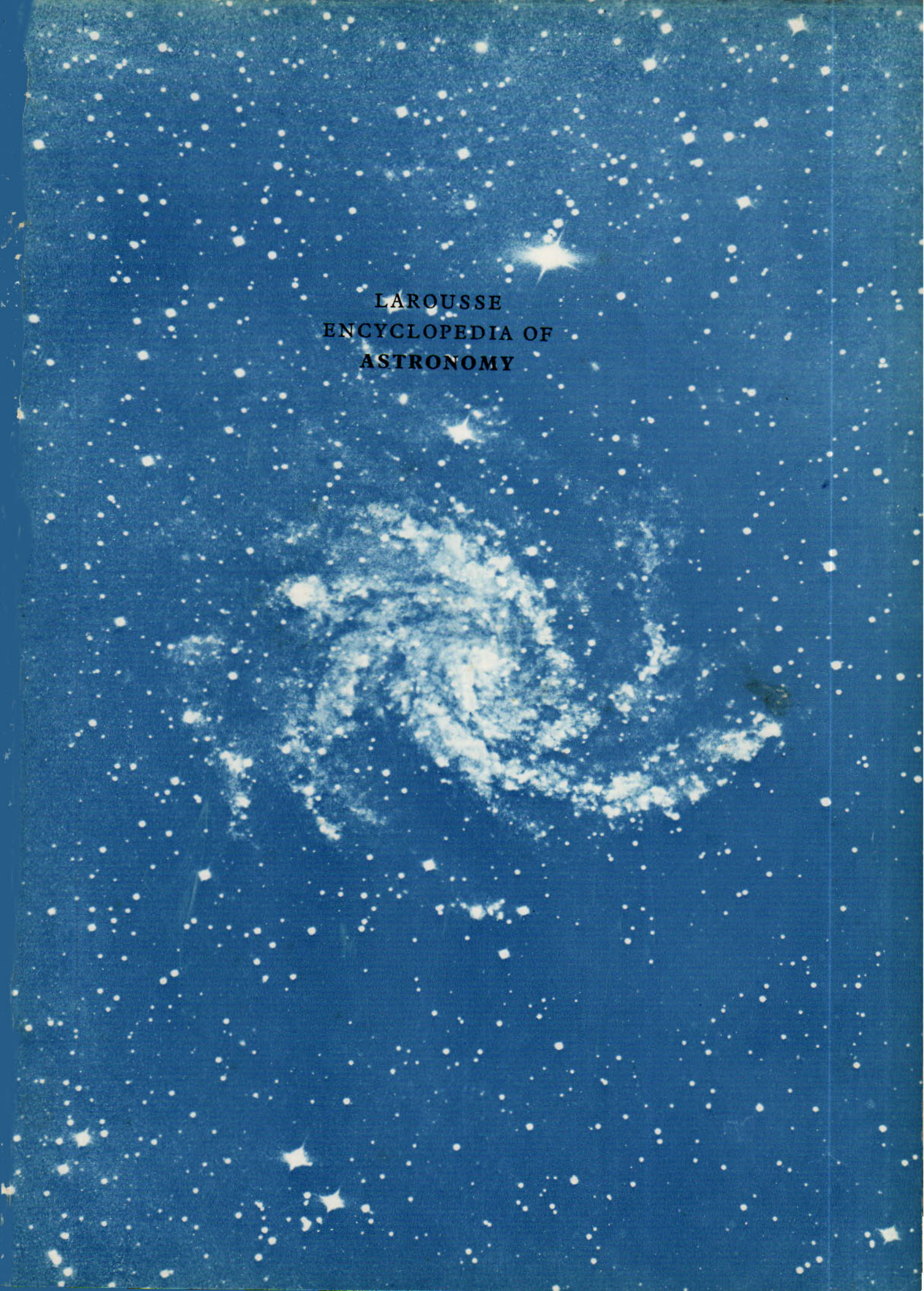
ASTRONOMY

ASTRONOMY





LAROUSSE ENCYCLOPEDIA OF ASTRONOMY

LAROUSSE ENCYCLOPEDIA OF ASTRONOMY

BY LUCIEN RUDAUX AND G. DE VAUCOULEURS

With an Introduction by
F. L. WHIPPLE
Professor of Astronomy, Harvard University

READERS UNION
PAUL HAMLYN
LONDON
1962

LAROUSSE ENCYCLOPEDIA OF ASTRONOMY

Translated by Michael Guest, F.R.A.S. and John B. Sidgwick
and revised by Z. Kopal, Professor of Astronomy
at the University of Manchester
from

LAROUSSE ASTRONOMIE

first published in France by
Augé, Gillon, Hollier-Larousse, Moreau et Cie.,
The Librairie Larousse

This copy is one of a small edition published by Readers' Union Limited, and made available to its members by arrangement with Paul Hamlyn Limited. Full particulars of Readers' Union are available from Readers' Union Limited, at Aldine House, 10–13 Bedford Street, London, W.C.2, and at Letchworth Garden City, Herts.

Printed in Czechoslovakia
Illustration positives made in France and England
© *Paul Hamlyn Limited 1959*

INTRODUCTION

by F. L. Whipple

WE ARE LIVING in the space age. Man has broken the tyrannical bounds of gravity that confined him so long to earth, only a tiny speck in the cosmos. Now man's machines and devices travel across the gravity barrier in the early exploration of a universe so vast that it defies comprehension. No longer are man's eyes and instruments blinded by the black curtain of atmosphere that shuts out most of the incoming messages of radiation from outer space. Now vehicles shot beyond this atmosphere contain instruments that can 'see' the universe by the ultra-violet light, x-rays, infra-red light and radio emanations that could never penetrate our cloudy, opaque and tremulous atmosphere. Even man himself has made his first voyages into space.

Man has developed a great insight concerning the sources of energy in the universe. Reactions in the nuclei of atoms that keep the stars burning for billions of years are now well enough understood and controlled to turn the wheels of industry. Although the stars have not shown man directly the secrets of their radiation we have learned from them that such secrets exist. The stars have served as beacons to give man the hope and assurance that by continuous delving into nature's treasury of knowledge he might expect to find answers to questions involving practical power sources as well as evolutionary secrets hidden for hundreds of millions of years.

Some might argue that for the first time in man's history astronomy has taken its rightful place, because problems of the nature, the past, the present and the future of the physical universe represent the greatest challenge to man except perhaps man himself. But the significance of astronomy in man's upward climb from the cave-men must not be discounted in the dark ages of the past. Every sunrise presented the miracle of the newly born sun in some older philosophies, while each sunset represented its death. The stars have reflected man's own concepts, aspirations, fears and hates in countless different forms in cultures of the past. Only rarely do the stars and planets stand for evil, the undesirable or the dreaded. Usually the heavenly bodies mirror the idealism and hopes and aspirations of man and reinforce his will towards life and progress.

The late Sir Arthur Eddington, the great British astrophysicist, believed that physicists on a cloud-bound planet could by sheer deduction conclude that stars were possible and indeed delineate their character. None will contest Eddington in this statement but one might seriously doubt whether intelligent beings living without the ostensible evidence for extra-terrestrial

objects could have risen far enough above the superstitious mumbo-jumbo surrounding infinity to have applied their physical knowledge in such deductions. The many thousands of years during which the heavenly bodies were primarily religious or superstitious symbols testify to man's weakness in cutting boldly with logic as a sword through the jungles of superstition. One may question seriously whether science and technology would develop at all in a cloud-bound planet where no direct signs of the outside universe were manifest except for the alternation of night and day. Without the sun, moon, stars and planets to prod his spirit, mankind might still consist of a few hundred thousand persons gnawing on bones in the warmer caves of the planet. Whatever the answer to this question, the astronomical development in our earth cultures has inevitably been a representative measure of the progress of the culture generally. Anthropologists use this criterion.

Some may mourn that in our modern space age accurate measurements and clearer vision are dispelling many of the mystic veils and charming imagery of our past ignorance. In contrast, I feel that truth is beauty and that an understanding of space, time and the physical universe can only lead to a more interesting and captivating view of the universe. It may even lead to man's better understanding of himself.

If you who read these lines do not share my conviction that knowledge and understanding of our universe will lead man forward, I could easily point out the everyday practical uses of astronomy and astrophysics. However, such a detailed exposition is hardly necessary as one watches the unfolding of transportation, communications and earth studies made beyond the earth's atmosphere. Space is now as much a part of our arena of technology as is the atmosphere, the oceans or even solid earth. The peculiar problems of life in zero gravity, the problems of operating in a vacuum, the hazards in space from meteoritic impact, the dreaded Van Allen belt of radiations only 600 miles above the earth's equator, the problems of rocket propulsion and the peculiar difficulties of returning to earth through the atmosphere at speeds of several miles per second, are all real and practical aspects of astronomy today. These and other problems of space technology demand a great increase in our astronomical knowledge and understanding. While it is true that a bird does not need to study aerodynamics, chemistry or optics to survive comfortably in our atmosphere, I must point out that it has taken a hundred million years of natural selection to produce this unconscious adaptation to practical survival. I, for one, hope that man can continue to progress at a much more rapid pace in his conquest of space. To do this he must utilize to the full his intelligence, which, to become fully effective, must operate from the firm foundation of knowledge and understanding accumulated in basic studies of the universe. If man is to exploit this heritage of our past and look forward to uncovering new wealth of understanding in the future, he must begin with study, continue with study and follow with contemplation and with experimentation.

The basic foundations of astronomy are presented in the pages of this book. The exposition is clear and the profuse illustrations make understanding both easy and desirable. The broad treatment of practically all phases of basic astronomy should provide a continuous challenge to the reader who is meeting for the first time the jewels of knowledge and understanding that have been so painstakingly uncovered over many centuries. A careful reading of these pages will constitute a platform of background understanding from which the reader can reach out from a firm foundation to investigate areas of peculiar interest to himself. If reading these pages does not create a spark of interest in further study of the universe, the reader will have the satisfaction of understanding far more clearly than before the true significance of explorations

of the universe made from astronomical observatories, from space satellites or from exploratory missions into the depths of our solar system. These accomplishments will then become significant realities to him and not merely reports of incomprehensible activities by 'geniuses', by 'mad scientists' or by military forces of dreaded import. These scientific triumphs will become understandable in terms of man's conquest over nature and his climb up the ladder of evolution towards goals that are yet undreamed of.

The reader will learn to appreciate that many of the theories based upon the best information available must be altered when new methods of investigation and new facts point to entirely different conclusions from the preliminary pointing. Thus he will find, perhaps to his surprise, that scientists change their minds and that the finely developed structure of a scientific edifice may look entirely different from what might have been anticipated from the ground plans. In a number of areas the presentation in the following pages does not represent the 'last word' of scientific development at the time of the manuscript's completion. This is particularly true in the area of evolution, both of the planetary system and the universe, in the area of interplanetary material at large, in the area of the moon and in the area of space science by rockets. In many fields the progress is enormously rapid and clarifications in older concepts are made almost daily. The measured age of the solar system seems to be continuously increasing while an individual's philosophic leaning towards a static or an exploding universe can almost equally well be satisfied by the observational data. On the other hand, simple matters of fact are changing in some areas where new methods of measurement and theory have altered the interpretation of older observations. Developments in our knowledge of the moon can be followed in the newspapers.

The co-operation of 53 nations in the International Geophysical Year 1957-8 represented the beginning of an enormous acceleration in studies of the earth and space. This impetus did not die at the end of the official 18 months of international co-operation on earth and space studies, but has continued. The increasing pace of scientific research and investigation makes nearly impossible the already difficult task of the popular presentation of scientific frontiers.

Nevertheless, the basic background, methodology and concept of astronomy change slowly and a huge array of knowledge that ties together so many facets of the universe is essentially understood. Great steps forward may be made, of course, in any direction at any time. However, the reader of this book can assure himself of a basic foundation in the subject from which he can then appreciate the changes and new developments which constitute the basic fascination of astronomy and other sciences. For, without the unknown, what would there be to explore?

CONTENTS

INTRODUCTION *page* v

BOOK I THE SPLENDOUR OF THE HEAVENS

1 Stars in space *page* 3
2 Our observatory, the Earth 23

BOOK II THE EMPIRE OF THE SUN

3 The Solar System 57
4 The Earth as a planet 72
5 The Moon 113
6 Eclipses, occultations and tides 155
7 The planets 177
8 Comets 229
9 Meteors and meteorites 245
10 The Sun 259
11 In space 290

BOOK III THE REALM OF THE STARS

12 The stars in the sky 301
13 Double and variable stars 312
14 The stars as suns 335
15 The nature of the stars 349
16 The galaxy 360
17 The extra-galactic nebulae and the expansion of the universe 395
18 Past and future 427

BOOK IV ASTRONOMICAL INSTRUMENTS AND TECHNIQUES

19 Instruments and methods of measurement 451
20 Spectroscopic analysis and its astronomical applications 478
21 The philosophical, scientific and technical value of astronomy 498

INDEX 500

COLOUR PLATES

		facing page
I	General map of the Moon	*144*
II	Total eclipse of the Moon	*166*
III	The planet Mars	*198*
IV	The planet Jupiter	*214*
V	The planet Saturn	*222*
VI	A solar prominence	*262*
VII	Stars of different magnitudes	*310*
VIII	The constellations of the northern hemisphere	*between pages 310 & 311*
IX	The constellations of the southern hemisphere	*between pages 310 & 311*
X	The constellations of the equatorial zone	*facing page 311*
XI	Coloured multiple stars	*326*
XII	Principles of spectroscopic analysis	*486*

BOOK I

THE SPLENDOUR OF THE HEAVENS

Fig. 1 — THE EARTH IN SPACE

Fig. 2 — Stars in the night sky. Twilight fading on the horizon

CHAPTER I

STARS IN SPACE

Appearance of the sky. — In the presence of the sky, everything at once enraptures our gaze and arouses our curiosity. Above us, moving in slow, unwavering circular fashion, there comes into our view the dazzling disc of the Sun, followed by the ever-changing shape of the Moon and the thousands and thousands of stars, twinkling points of light appearing in turn after the close of day. Then there are the more unusual phenomena, strange-looking comets, eclipses of the Sun and Moon and the sudden fleeting appearance of a bright meteor. So many splendid sights cannot fail to leave a vivid and profound impression on us, and it was from this visual attraction, right from the earliest days of conscious man, from his sometimes wondering, sometimes trembling contemplation of the heavens, that astronomy was born. It is perhaps the most exciting of all branches of science.

Astronomical knowledge was founded solely and simply on what the human eye could see. Although this means of investigation works efficiently in many cases, it is sometimes liable to lead us into error. In the first place it produces one illusion which only reasoned thought can destroy: that of the 'celestial vault' whereby all heavenly objects seem to be placed side by side, like lights on the inside of a spherical dome, appearing however at some indeterminate height. Such an impression is explained when we realise that our eyes are unable to detect anything more than small differences in distance by any visual impression of depth, and that they arrange all the stars in the same plane however enormous and unequal may be the distances really separating them.

But, false though it may be, the 'celestial vault' idea persists, and it is not only used in star charts and in astronomical diagrams generally, but also figures in the design of celestial globes and planetaria. In short, we think of the sky as a sphere of infinite radius at whose centre we ourselves are placed.

So when we speak of the 'sky', we must understand it to mean the general spectacle which is unfolded there, not the real distribution of stars in space which constitutes it, this being due to the limits of our vision and effects of perspective. Thus the 'starry sky' is but an image. No doubt this image is suggestive, but it has not the same reality as when it is used, as it normally and indiscriminately is, to designate the heavenly domain. In fact the atmosphere in which we are immersed, and which is indispensable to the maintenance of life, is a layer of gas which extends above the earth and follows the curvature of the globe. The vaulted impression which it produces is often made physically visible by layers of cloud at different heights which combine to form a veritable ceiling sloping away all around us to the horizon. The state of the atmosphere also determines how well we see the stars, which must of course be observed through the more or less transparent blanket of air interposed between them and our eyes.

The meteorologist's sky, with its classical azure tints,

is thus very different from the astronomer's, and we must not confuse the two in the way of those people who think that everything that goes on above their heads is all part of one and the same thing. We shall see, too, what part the sky plays in the production of certain astronomical phenomena, the most important from the practical point of view being the 'day' and the conditions under which we enjoy the Sun's light.

People have very frequently compared the astronomical sky, whose successive aspects are so much a visible translation of the universe, with the pages of a book open for all to see. But we must know how to read this book, for to many it is written in an unknown language. However, while demanding a lot of hard study, it rewards us with a profound joy as soon as we begin to master it, first spelling it out, then understanding what it is about, seeking to discover the secrets it contains. In the end the inquiring mind will find it far more interesting than any novel. But, it may be said, you are comparing a story, a series of dramatic events, with the spectacle of the heavens, which, despite their imposing splendour, always remain essentially the same, a few periodic variations excepted. Here again, however, appearances are deceptive. For to seem changeless during the brief time we have to watch, with what tremendous, never-ending, never-hesitating movement must the universe be filled! Our imagination is bewildered by the immensities of space and time that we discover, and we are astonished by the transformations and cataclysms that are everywhere revealed by modern science. If men cannot take in everything at first glance, it is because they are too small on the scale of such happenings, because their senses are imperfect, their lives too brief. For the seeker, however, such knowledge comes to be his most precious possession. But we cannot completely understand what is happening unless we also know this: the Earth on which we live is a heavenly body like any other. It is plunged like the rest in the depths of space and shares their movement. What we actually see is a combination of our own movement and that of all the bodies which we can see in every part of the universe.

Thus, despite all the pleasure to be gained from it, we cannot be content merely to look at the sky. We must sort things out, see what happens, how and why such things appear and such phenomena occur at any given moment. Therefore, on the one hand, we must be able to recognise all the different sorts of stars, for although they may all look the same their real characteristics differ considerably; on the other, knowing what they are will then help us to understand the movements they undergo and the changes which appear from time to time. Lastly — and this is the goal at which our human curiosity is directed — we wish to find out if, among these at present inaccessible worlds, which seem so reluctant to divulge their secrets, there may not be one or two like our own, inhabited by beings more or less resembling ourselves.

It will easily be understood that the diversity of these problems means that they must be attacked in different ways. In fact, we cannot be content merely to gaze at the stars, simply to examine them by means of telescopes and cameras. That would be to remain in the sphere of descriptive astronomy, extremely seductive in many ways but inevitably failing to give us the complete understanding of the universe that we seek. To this branch of astronomy must be added, for example, the observation and measurement of positions, and these, when analysed by means of the appropriate mathematical formulae, enable us to define the nature and magnitude of the movements involved; from this we can then define the laws which they obey. But essential as these data may be, we cannot rest there. By every possible means we must attempt to determine the nature of the heavenly bodies, their chemical composition, the character and origin of the phenomena which take place within them. Relatively speaking, this branch of astronomy — astrophysics — is quite recent, and its tools and techniques are still in process of development, but it has already made discoveries to which the human eye unaided could never have aspired. Later we shall see that the tremendously important results achieved by this new science have completely overthrown earlier notions; but it must be said at the outset that this sort of investigation relies on means whose application stems from the principles and theories of

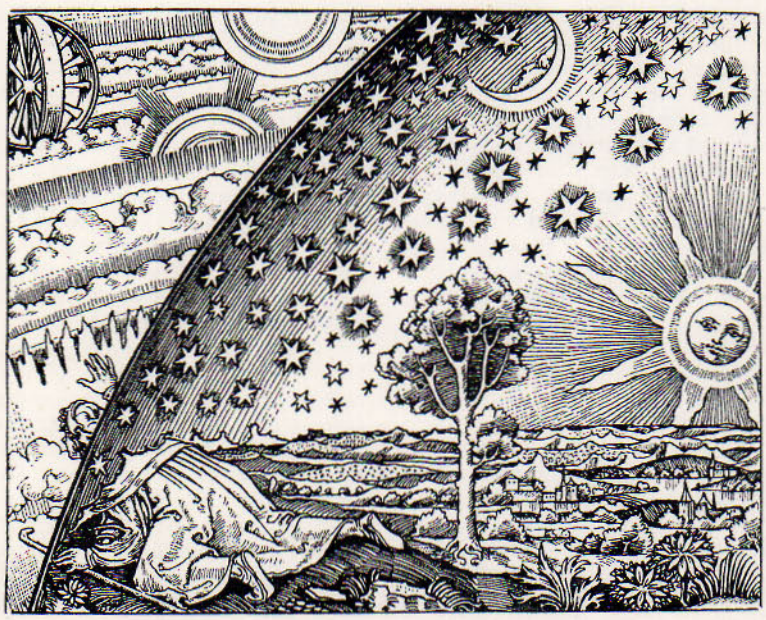


Fig. 3 — THE SYSTEM OF THE WORLD ACCORDING TO THE IDEAS PREVALENT IN THE MIDDLE AGES. Above the flat Earth the sky forms a vault studded with stars. An inquisitive traveller puts his head through the vault and discovers the complicated mechanism which moves the stars.

Fig. 4 — SHAFTS OF SUNLIGHT DIFFUSED BY DUST SUSPENDED IN THE ATMOSPHERE

Fig. 5 — A 'CEILING' OF CLOUD GIVING THE ILLUSION OF A VAULT TOWERING ABOVE US

modern physics, which are often very difficult to expound simply. Therefore some of our explanations may perhaps seem laboured at times, although we shall do our best to avoid this as much as possible.

Different types of heavenly bodies. — In descriptions the following terms are used: Sun, Moon, Planets, Satellites, Comets, Shooting Stars, Stars, Nebulae. Unfortunately some of these have a double usage which leads to confusion.

The Sun is the most striking of all heavenly bodies. Our eyes cannot tolerate its brilliance when it beats down from high in a clear sky, but when it is veiled by semi-transparent clouds, or its light is dimmed by the increasing absorption of the atmosphere near the horizon, then it can be seen as a perfectly circular disc (except where, as we shall see, refraction in the Earth's atmosphere has the effect of distorting it into strange shapes). This flat disc appearance is the result of seeing a globe which is too far away for its spherical nature to be apparent through any relief effect. The same thing happens in the case of the planets, which are likewise spherical and, owing to their distances, very reduced in size. As for the rays which are nearly always shown round the sun in diagrams and pictures — they are non-existent, being merely an attempt to portray certain optical effects produced by our eyes whenever we look at a very brilliant source of light. On the other hand the oft-used expression 'solar rays' should be taken to mean 'radiation', the stream of light- and heat-waves which illuminate and warm anything exposed to them. Similarly we refer to the rays of the Moon and stars. Finally, much as we may admire the

Fig. 6 — APPARENT MOVEMENT OF THE SKY. The curved lines on this photograph, taken with a time exposure and a fixed camera, show the movement of the stars in the sky.

Fig. — 7 THE SUN'S DISC SEEN THROUGH LIGHT CLOUD

spectacle they produce, some are not what we may properly call rays but atmospheric effects of light and shade caused by means of sunlight filtering down through gaps in the clouds, made visible by their encounter with dust particles in the air.

The Sun gives out its own heat and light like a tremendous furnace with an intensity far exceeding anything we know on Earth. The reason we are not roasted to death by this fire is that it is so far away. For the moment, however, let us think of the Sun merely as a great lamp by whose light we are able to observe the planets.

Under this heading we must consider a large number of worlds composed of materials of different densities, which are not self-luminous. We are able to see these bodies solely because their surfaces are illuminated by the Sun. They will appear brighter the nearer they are to the Sun and also the lighter their surface colouring. Against the blackness of space these bodies seem very bright; we have the same impression when we look out at a brilliantly lit landscape through tiny gaps in the shutters of a darkened room.

The planets revolve around the Sun; the Earth, our home, is a planet.

The Moon is the best known of all heavenly bodies after the Sun, and it will serve as a good illustration of

Fig. 8 — SUCCESSIVE PHASES OF THE MOON WITH RESPECT TO THE SUN AND THE OBSERVER'S HORIZON

how the other members of the Solar System receive their light. The Moon circles the Earth and is our closest neighbour in space. Her phases attract the attention even of people who know nothing about astronomy. The Moon and the Sun are the only two bodies whose shape we can distinguish without the use of instruments.

The Sun shines only on that half of a planet which is turned sunwards. Therefore the Moon shows phases, that is, she appears to be illuminated from one side or the other, or face-on, according to her position with regard to the Earth and Sun. In a later section we shall discuss this more fully.

Being a solid globe and lacking intrinsic luminosity, the Moon is as much a planet as the Earth and the other members of the Solar System. But the planets appear only as points of light because they are at much greater distances than the Moon. For this reason the planets are often confused with the stars, a name indiscriminately applied to any bright point of light in the sky. Later we shall see just what the stars are, but even without any precise knowledge a little thought will enable us to draw the proper distinction. The real stars always remain in the same relative positions, which fall into the patterns we call the constellations — the Great Bear, for example, which everyone taking a stroll on a fine evening likes to be able to pick out. On the other hand the planets can be recognised, in spite of their stellar appearance, firstly because they are in general brighter than most stars, and

Fig. 10 — MOVEMENT OF A PLANET AGAINST THE BACKGROUND OF THE STARS

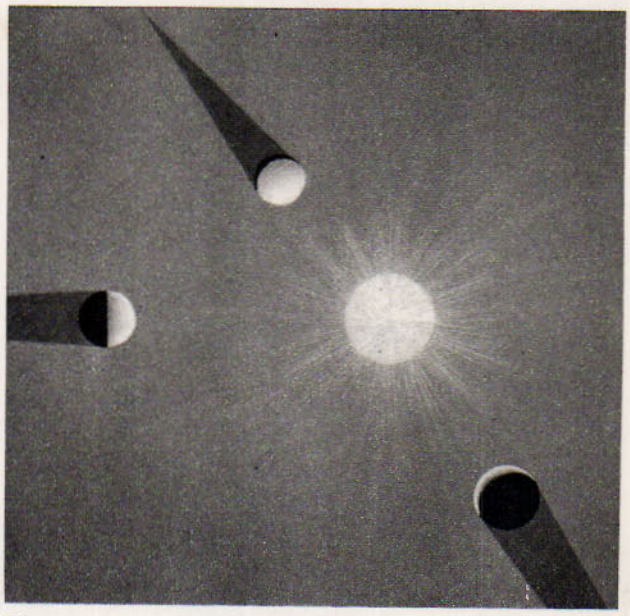


Fig. 9 — ILLUMINATION OF PLANETARY GLOBES BY THE SUN

secondly because they are constantly changing their position with respect to the 'fixed' stars. This is a consequence both of their proper motion round the Sun, and of their proximity to the Earth compared with the stars which are so far away that, for all practical purposes, we can think of them as pinpoints of light on a backcloth placed at infinity.

The planets are quite diverse in size and appearance; this is partly due to real differences and partly to their position when we observe them. But these are things that the human eye alone cannot see for itself and which can only be shown with the help of powerful telescopes.

Some of the planets are accompanied by smaller bodies called satellites, which circle their own particular planet just as the Moon circles the Earth. The Moon, then, is the satellite of the Earth, and its only natural one.

Comets used often to be called 'hairy stars'. Their appearance is characterised by a bright streak called the 'tail' emanating from a luminous area called the 'head'. The composition of comets is not yet fully understood. Their nebulous light is partly intrinsic and partly reflected

Fig. 11 — A PLANET, JUPITER, AND ITS SATELLITES

from the Sun, around which they describe orbits in the same way as the planets, according to the same universal laws. When we make our detailed study of comets, whose appearance has so often been greeted with superstitious fear, we shall see that the description just given is not quite accurate, in so far as certain comets show up merely as a faint patch of luminosity without any sign of a tail. For the most part they are dim objects which would pass unnoticed but for being picked up in our instruments.

The name 'shooting stars' is descriptive but totally erroneous. Contrary to popular belief, they are definitely not stars which have come 'unstuck' from the 'celestial vault' and started falling towards the ground. The sudden appearance of an exceptionally bright meteor, or 'bolide', or a rapid succession of shooting stars giving the impression of a shower of luminous rain, can produce a spectacularly beautiful effect, but these phenomena originate within the confines of the Earth's atmosphere, not in the depths of space. In fact shooting stars are the result of the Earth's encounter with small particles of matter moving round the Sun.

Having surveyed the general characteristics of the various types of heavenly bodies, let us now turn from the members of the Solar System to the stars themselves, and see how they differ.

Fig. 12 — A COMET

The stars are suns more or less comparable with our own, that is, enormous spheres of glowing gas whose emission of light and heat is such as to beggar the imagination. But these suns are so far away that however large and bright they may really be, none looks any bigger than a point of light, no longer dazzling the eye, while

Fig. 13 — A 'SHOWER' OF SHOOTING STARS

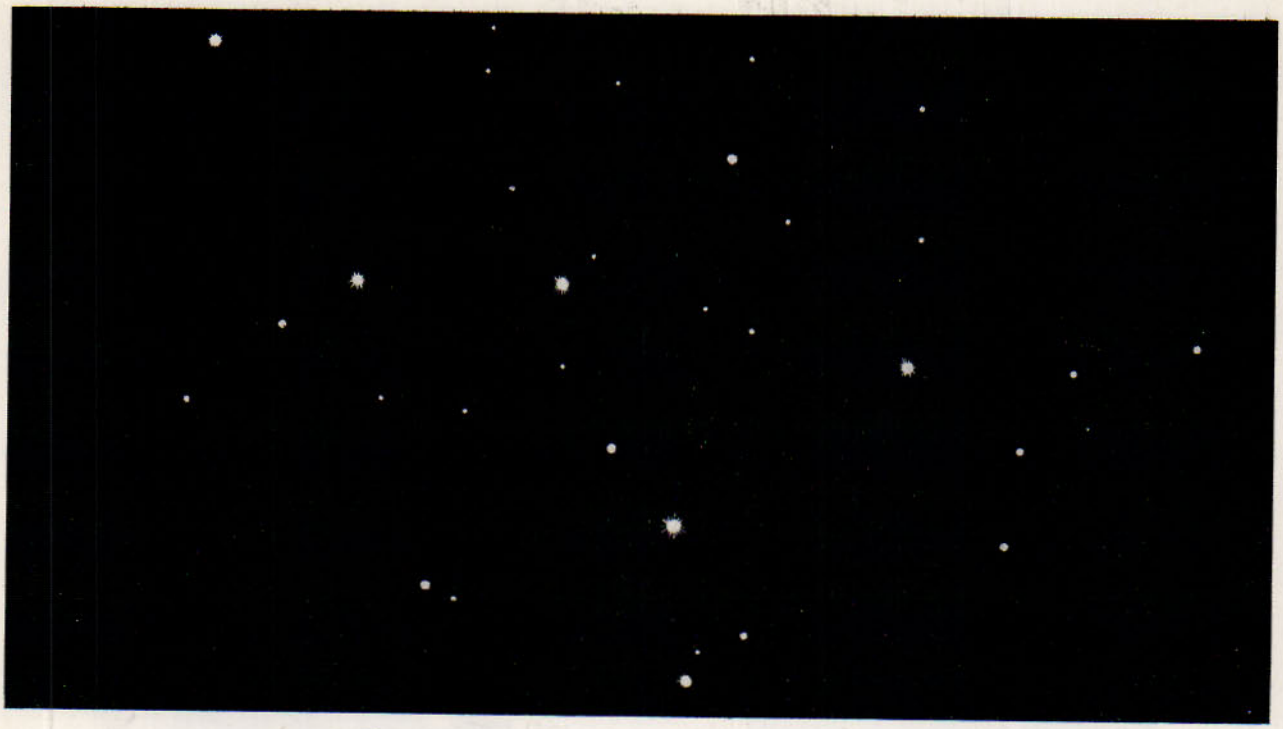


Fig. 14 — PART OF THE SKY SEEN WITH THE NAKED EYE

the vast majority are no more than the feeblest gleams lost in the depths of space.

Observe the stars, whose distribution is responsible for possibly the best-known spectacle in the sky: even the most cursory glance will reveal that they are not all of the same apparent brightness, since they range from the very brilliant to those on the point of invisibility. But the stars visible to the naked eye represent only a very small proportion of the heavenly multitudes. Beyond the range of the naked eye, our modern optical instruments have revealed the existence of such huge numbers of faint stars that where with the unaided eye we can hardly pick out more than the astronomers of old did, the telescope shows millions and millions. No words can do

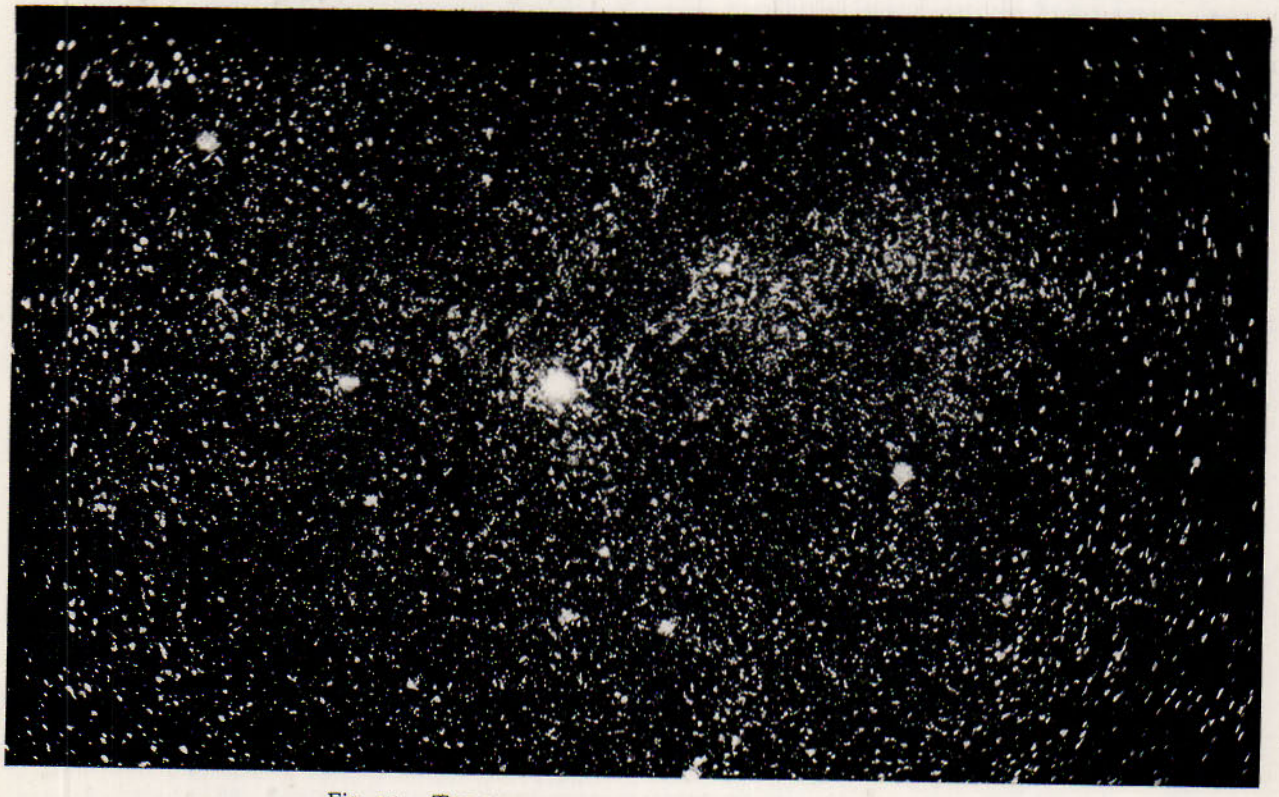


Fig. 15 — THE SAME REGION SEEN THROUGH THE TELESCOPE

them justice. Only the realisation of the facts can bring home the grandeur and richness of such a spectacle. Compare Fig. 14 and Fig. 15, showing the same part of the sky, one as we normally see it, the other a photograph taken with a moderate exposure.

Having once been introduced to this incredibly dense population of stars seeming to fill space with luminous dust, we can immediately explain other phenomena. The first of these is the Milky Way, a sight which we all admire on a fine summer's evening. Girdling the Earth like a vaporous sash, irregular in outline, uneven in brightness, the Milky Way is the result of a tremendous crowding together of extremely faint stars, individually invisible to the naked eye, but whose light, when seen merged together as a whole, takes the form of this nebulous ring of luminosity.

Fig. 16 — A SPIRAL NEBULA

Fig. 17 — A GASEOUS NEBULA

But the sky holds yet more surprises than these. Sometimes where nothing was seen before, we witness the sudden and unexpected appearance of a star destined perhaps to outshine all the others in the sky before falling back into obscurity. At one time the bursting of a nova, as it is called, would be greeted by the populace with much trepidation, but nowadays it merely provides astronomers with a great many interesting problems.

Lastly, almost lost amidst the multitudes of stars, but for the most part much more distant, are the nebulae. These are patches of pale light, some regular, some irregular in shape, some in the form of a spiral like a Catherine wheel. But although the designation 'nebulae' was adopted at the time of their discovery with instruments just powerful enough to reveal them, and corresponds with their superficial appearance, we must nowadays distinguish differences in type, for these objects are by no means all similar. First there are the real nebulae, that is, clouds of luminous gas or sometimes dust; secondly there are the nebulae resulting from the agglomeration of faint stars, each too faint to be visible separately, as we have shown in the case of the Milky Way seen with the naked eye. Generally speaking, the second type of nebula exhibits a spiral structure of greater or lesser development.

The movement of the stars. — When we consider the tireless, regular way the stars seem to move, we might be tempted to believe that the heavens are presented to us as a symbol of unity and concerted harmony. But then we notice that from one night to the next the Moon's position has changed, and that this change can be seen in a matter of hours if she happens to be near some bright star that we can use as a reference point. In the same way, the position of the planets is constantly varying against the background of the constellations, which, however, always remain the same.

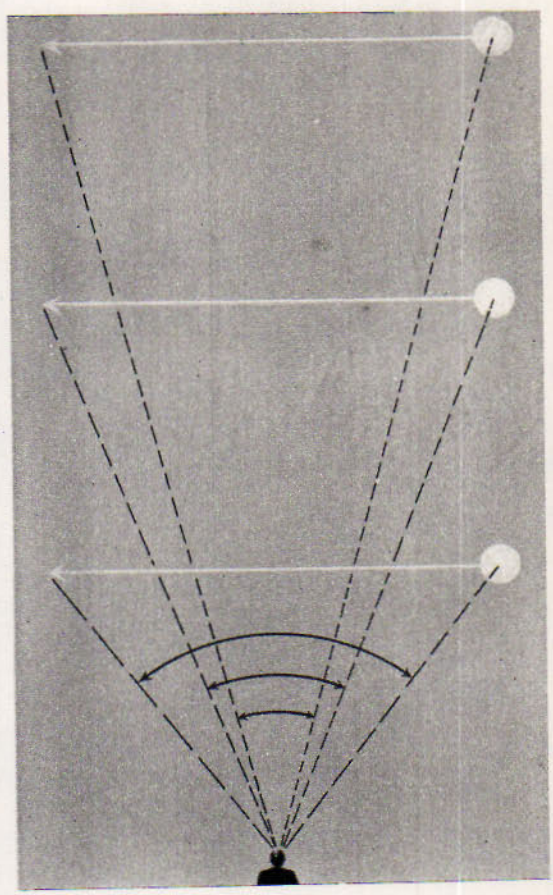


Fig. 18 — VARIATION OF APPARENT ANGULAR VELOCITY AS A FUNCTION OF AN OBJECT'S DISTANCE

Fig. 19 — THE MILKY WAY AS SEEN WITH THE NAKED EYE

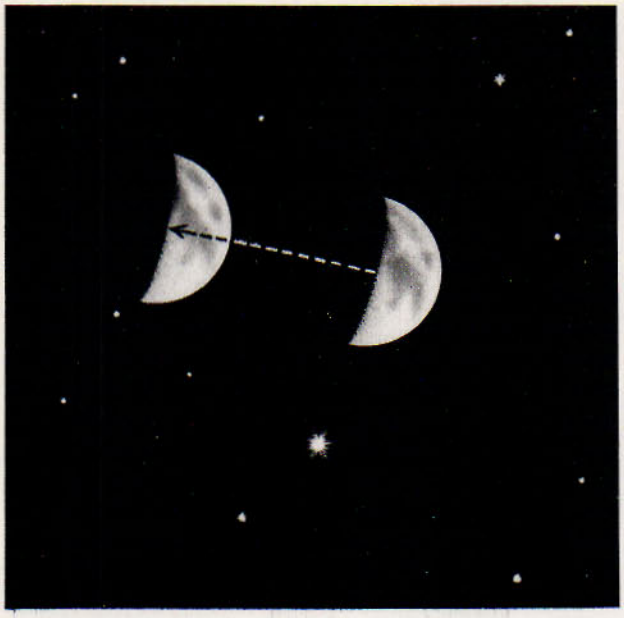


Fig. 20 — MOVEMENT OF THE MOON

These few examples give us a first indication of movement occurring in space. Now we must generalise. In fact not only is nothing fixed, but some stars are moving with considerable speed; it is just that we cannot appreciate these movements for what they really are on account of the tremendous distances involved. Moreover, in our daily lives we are accustomed to receiving constant proof of movement. In a car, we appear to see the surrounding landscape moving past: the nearest objects travel a considerable distance in, say, one second, whilst with more remote objects the change in position in a given time becomes less marked until at the horizon it becomes nil. This explanation will recall to mind certain of the laws of perspective. In space we find many bodies moving like projectiles, each with its own particular velocity, all at widely different distances from the Earth. The Moon, for example, is very close to our world, and in order to accomplish her monthly journey round us she must travel at about 1 km. per sec. On the other hand, planetary velocities range from 25 to 50 km. per second — whilst the stars are moving at least as fast and in most cases very much faster, with velocities up to 300 km. per sec. Lastly it is known that the spiral nebulae or galaxies, whole stellar systems which are the most distant objects we recognise today, are moving as a whole at the most incredible speeds, from a few hundred up to 60,000 km. per sec.

Now what do we actually see? We see the Moon, with its relatively modest speed that even a man-made projectile could rival, moving visibly against the sky from hour to hour, whilst the planets, with their greater speeds but greater distances, take much longer to move the same amount, and the stars seem permanently fixed. During the centuries that men have watched the skies, the relative positions of the stars which mark out the constellations have not undergone any change perceptible to the naked eye. Only the most accurate and painstaking measurements, made over the course of many years, can detect these movements.

If we really grasp these basic ideas when we contemplate the universe around us, we shall be in a better position to understand this diversity of movement and the effects it produces. When, for example, we see the Sun begin to disappear in the full light of day as if being eaten away by some unknown agent, we can remember that this phenomenon, which we call an eclipse, is simply the result of the Moon's moving in front of the Sun's disc like a black screen. When the Moon undergoes a similar fate, it is simply that she has moved into the Earth's shadow, so cutting herself off from the source of light by which she shines. Primitive peoples view eclipses with fear and trembling: we do not need to.

At certain times we may witness the beautiful spectacle of a conjunction, that is a close approach between two heavenly bodies — the Moon and a planet, for example, or a planet and a star. In fact their approach can be so close that one of the two bodies will be eclipsed by the other: this is called an occultation. Now when a conjunction brings two objects close together on the celestial sphere like this, we have an ideal opportunity to observe their relative movement. We notice that the eye is unable to tell which of the two is the nearer until one occults the other; even then, if the approach is between a planet and a star, we still need the aid of a telescope. But eclipses and occultations are not simply beautiful spectacles: they are also important sources of information about conditions on the Sun and planets, as will be shown when we examine the question of the solar corona and the possibility of a lunar atmosphere.

Types of movement in space; orbits. — Nothing in space moves in the rectilinear fashion of a projectile fired along an unswerving path at uniform velocity. All celestial movements take place in geometrical curves (or orbits) of various dimensions, the speed varying according to certain universal laws which we shall come to later.

Fig. 21 — CONJUNCTION OF TWO PLANETS

Fig. 22 — An eclipse of the sun

Fig. 23 — An eclipse of the moon

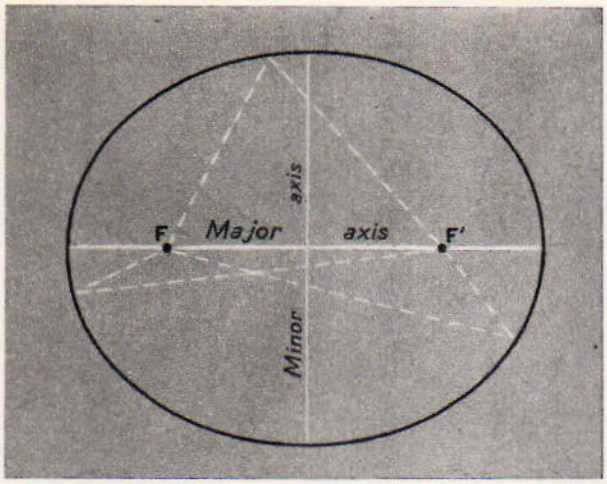


Fig. 24 — PROPERTIES OF THE ELLIPSE, F, F', FOCI

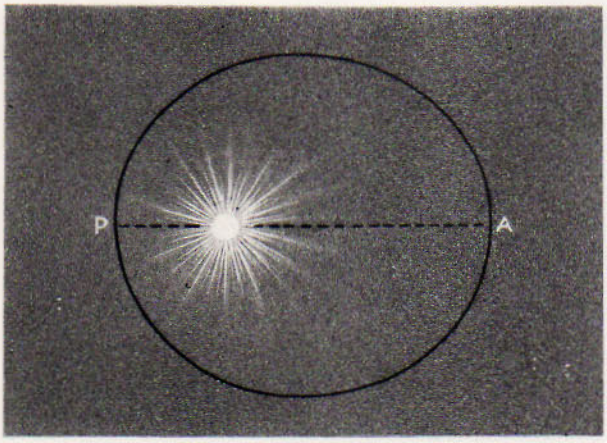


Fig. 25 — MAJOR AXIS OF AN ORBIT. P. PERIHELION; A, APHELION

For the present it is enough to say that every orbit is centred round a heavenly body (or group of bodies) whose greater attraction dominates the movements of the smaller bodies. In company with the other planets, the Earth goes round the enormous central Sun, whilst the Moon goes round its primary, the Earth. And when we come to the study of the stars, we shall find that they too are subject to similar forces. The general movement of the stars forming our galaxy presents a complex problem which, owing to the apparent minuteness of the movements in question, is only now being satisfactorily resolved.

As a general illustration, we can compare an orbit with a circular race track, where the runner comes round to the same point again and again after an interval determined by this speed and the condition of the track he has to cover. The time taken by a planet to complete one circuit is called the period of revolution.

Orbits are ellipses, the ellipse being a geometrical figure drawn round two fixed points called foci. The primary body, the Sun in the case of the Solar System, occupies one of the foci. The sum of the distances between the foci and any point on the curve is constant, this sum being equal to the greatest diameter of the ellipse, the major axis. This property is often used by gardeners to mark out oval flower beds: two stakes are driven into the ground at points corresponding to the foci, and a cord is attached to them by each end. The length of this cord is made equal to the length of the major axis required. The cord is then stretched round a spike held in the hand, and when this is moved round, keeping the cord taut, the ellipse is automatically marked out. The greater the distance separating the foci in proportion to the length of the major axis, the more oblate will be the ellipse produced: the relation between the inter-focal distance and the length of the axis is called the eccentricity of the ellipse.

Orbits may be of any eccentricity. Some are almost perfectly circular whilst others are extremely elongated ovals. Comets, for example, have orbits which are so elongated as to approach the parabola or hyperbola. The latter are geometrical figures which may be simply defined as ellipses of infinite axis. We can imagine them as ellipses at one end, the curve opening up and never closing again no matter how far the arms of the curve be extended. When we come to study comets, we shall see the consequences of this fact.

It can be seen that a body travelling along an elliptical orbit must at different times find itself at different distances from the large body occupying one of the foci; it will be nearest when at the end of the axis near the main focus, and furthest at the opposite one. These two points are called perihelion and aphelion respectively in the case of an orbit round the Sun (from the Greek *helios* — Sun); in the case of the Moon's orbit round the Earth, the respective designations become perigee and apogee (Greek *gê* — Earth).

Mass and attraction. — A force exists which keeps the heavenly bodies in their orbits and governs their movements: this force is called 'gravitation' or universal attraction, the discovery of which immortalised the name of Isaac Newton. To state this universal law is comparatively simple; but to understand it properly, we must first grasp one vital, indispensable idea, that of a body's mass.

It is very easy to confuse mass and weight. The mass of a body is the quantity of matter it contains, whatever

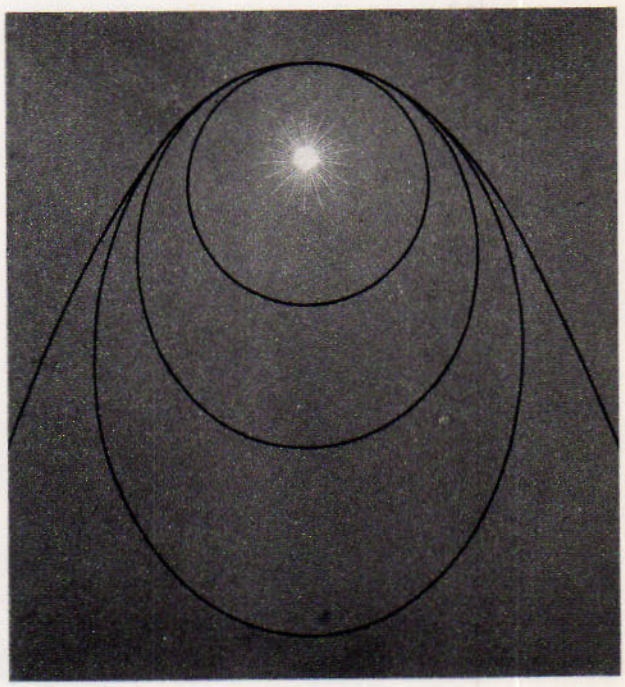


Fig. 26 — CIRCLE, ELLIPSES, PARABOLA

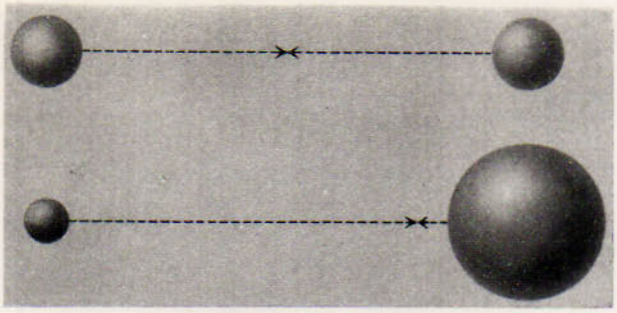


Fig. 27 — ATTRACTION BETWEEN BODIES

that matter may be, platinum or lead, rock or wood; whilst its weight, as we shall see in the chapter on the Earth, is simply a function of the force exerted on a body by the mass of a given body in the given circumstances.

Let us consider any two masses separated by a certain distance: everything happens as though an attractive force were acting between them. Newton defined the law which bears his name as follows:

The force acting between two bodies is directly proportional to the masses of the two bodies and inversely proportional to the square of the distance between them.

refute this theory, since we can put two objects on a table without their being drawn together by some attractive force. This is because in this case the attraction, which is extremely small, is resisted and overcome by the frictional forces between the objects and the table. In short, we are not dealing with the ideal conditions postulated above: we are on the Earth, whose own gravitational attraction enormously outweighs the forces exerted by any object that we can experiment with. Nevertheless, this does not invalidate the law, which can be verified in the laboratory by means of extremely sensitive measuring devices used with objects arranged so as to circumvent, if not the Earth's gravity, at least the friction produced by contact with a support.

Being isolated in what is conventionally called empty space, the stars can influence each other freely. They do not fall into one another, however, because they are endowed with movement; their speed of approach is governed by the attraction they undergo so as to maintain a constant equilibrium.

Let us explain these facts in the most elementary way possible by considering the behaviour of a projectile fired from a gun. Fired horizontally, it would travel in a straight line if its weight due to the Earth's gravity did not intervene. But under the influence of this attraction, which tries to pull it to the ground as it shoots along its path,

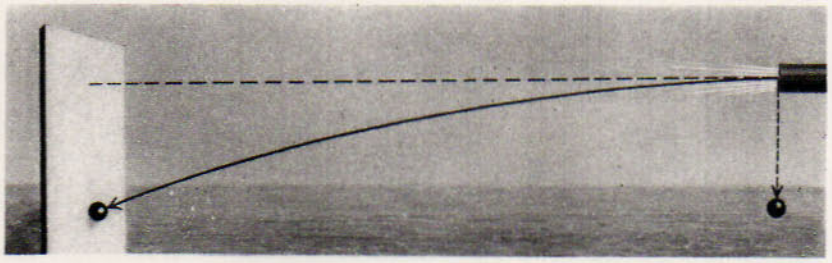


Fig. 28 — EFFECT OF GRAVITY ON A PROJECTILE

Let us imagine two large astronomical bodies alone in an otherwise empty universe. They may be any distance apart. Their mutual attraction immediately coming into play, they begin to draw towards each other with constant acceleration. If they are of equal mass, they will meet exactly at the half-way point. If they are of very unequal mass, the distance that each will have to travel will be proportional to their difference in mass, and the extremely large body will hardly move at all whilst the small one is drawn towards it.

It may be objected that everyday experience seems to

the projectile will follow a curved trajectory, the resultant of two component forces acting in different directions. It will not follow the straight line extending the initial direction of fire, but will fall earthwards in a given time by a distance equal to that through which it would have fallen vertically if simply dropped and not propelled from its point of departure.

The planets behave in an analogous fashion. Their orbits are trajectories constantly being bent towards the centre of attraction to which each is subordinate. The nearer they are to the centre, the greater the distance

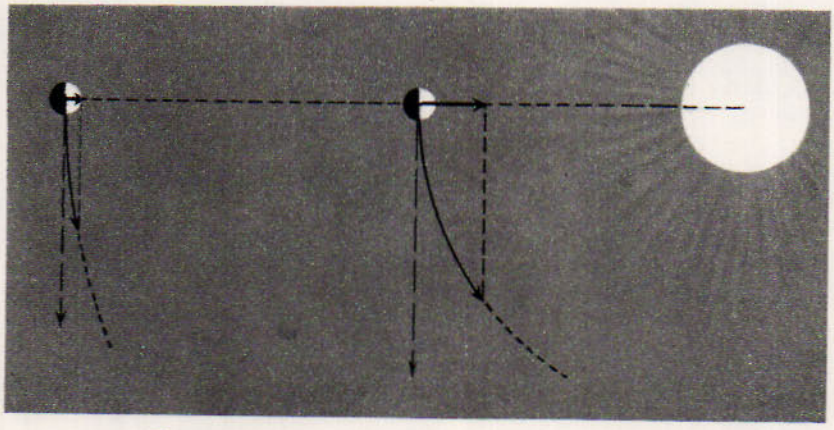


Fig. 29 — ATTRACTION AS A FUNCTION OF DISTANCE

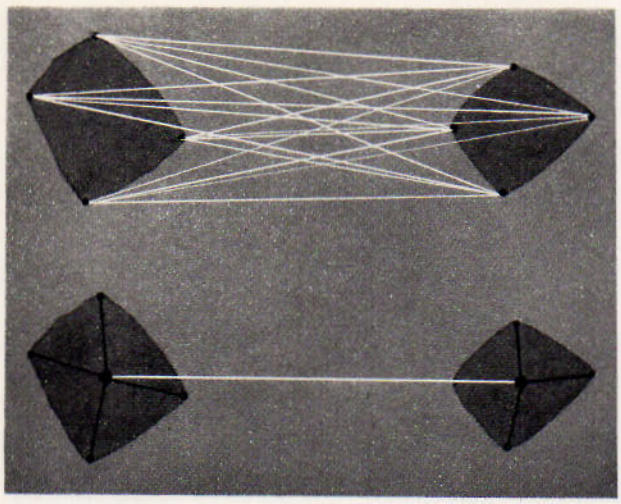


Fig. 30 – MUTUAL ATTRACTIONS AND CENTRES OF GRAVITY

they must fall towards it in one second. As in the case of the projectile, the gap between the straight line, or tangential direction, and the actual path followed increases away from the point of departure. The more the gap widens, the greater becomes the length of the curve, and since this must be covered in the same given time, it follows that the velocity must increase in proportion. Consequently, being at different distances, the planets must accomplish their respective journeys round the Sun at different speeds, the furthest being the slowest.

The centre of gravity. – The above explanation of the forces acting between two bodies may prompt one to remark, when going more deeply into the question: 'These two bodies, you say, attract one another with a force that varies as the distance. Now, if we consider that each body must have some definite volume, the distances separating various points in the one from various points in the other will not be the same. For example, points on sides facing each other will be nearer than those on opposite sides. Taking this argument to its logical conclusion, we can say that any molecule in the first body unequally attracts every molecule in the second, and vice versa. Therefore there are forces of variable intensity acting between various points in these two bodies. These primary, multiple attractions act along the straight line joining the points in question, and if we drew a diagram illustrating this effect, it would be a tangled network made up of an infinite number of lines.

How then may we reconcile this fact with the general exposition of the laws of gravity in which we treated the attracting masses as if reduced to geometrical points?

Fortunately the science of mechanics will help us to answer this question. It shows that in fact the overall attraction resulting from the individual attractions between all the molecules will act along a straight line precisely as if the whole mass of the body were concentrated at one particular point. This point is the centre of the primary attractions. If the bodies in question are relatively near together compared with their own dimensions, the positions of these points within their own object change with any alteration in their orientation or distance apart; this is easily understood when it is remembered that the attraction of molecule for molecule varies in inverse proportion to the square of the distance and that the relative distance of each point from any other point changes with any alteration of the distance separating the two bodies in question.

But if this distance becomes very great compared with the dimensions of the two bodies, then all the separate forces become effectively parallel, and the centre within each body is fixed at a position which may be considered invariable for all practical purposes. This is called the centre of gravity.

When the two bodies are spherical and homogeneous, the centre of gravity concides with the geometrical centre, and the system behaves as if the whole mass were concentrated at the centre. This is precisely the case with the stars and planets. As we shall see, their respective distances are enormous compared with the dimensions of their globes. Thus we can simplify the calculation of their gravitational effects by assuming that their mass acts as if concentrated at the centre.

However, this simplification is no longer permissible if the distance separating the two bodies is small compared with their dimensions; this occurs in the Earth-Moon system. Since the Earth is not absolutely spherical and since the Moon's orbit is obliquely orientated to ours and for the reasons mentioned above, the mutual attractions between the two bodies are constantly varying, thus bringing about certain irregularities in their orbital movement. From a study of the Moon's movements coupled with geophysical measurements, we can calculate the flattening of the Earth's globe. Moreover, the Moon pulls on the Earth's equatorial bulge, causing further variation in its movement.

Kepler's laws. The laws of universal attraction tell us what makes the planets go round the Sun.

Let us now consider the laws which govern the nature and speed of these movements. These laws, three in number, we owe to the genius of Kepler, who formulated them at the beginning of the seventeenth century. It was from studying them that Newton was led to formulate the more general principle of universal attraction or gravitation.

First law: The planets move round the Sun in ellipses, the Sun occupying one focus of the ellipse.

Second law: The radius vector connecting the Sun to a planet sweeps out equal areas in equal times.

Third law: The square of the revolution period is proportional to the cube of the major axis of the orbit.

There is no need to elaborate the definition of an ellipse

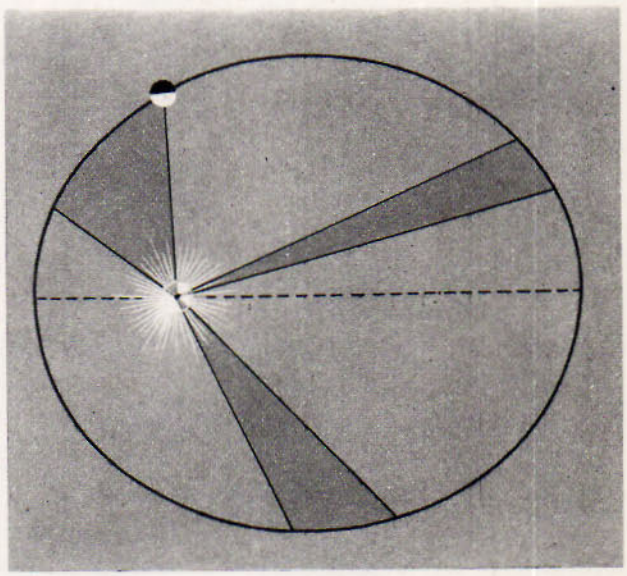


Fig. 31 – KEPLER'S LAW OF AREAS

given already. From the geometrical properties of the ellipse, it will be seen that the distance of a planet from the Sun must vary continuously at every point round its orbit. Important consequences may follow from this.

The second law connects what we have said about the effects of attraction and dissimilar speeds as a function of distance. Fig. 31 illustrates this principle. The radius vector, the straight line joining the Sun and planet, sweeps out equal areas in equal times. Owing to the eccentricity of the ellipse, equal surface areas must embrace unequally long arcs of the orbit. Now since it takes equal times to cover these unequal distances, the planet's velocity will be greater at the end of its orbit nearest the Sun than at the other. Thus we can see the importance of Kepler's second law, which governs the speed of a planet in its orbit: this variation according to distance has many repercussions, for example in the progress of the seasons on a planet.

The third law, involving a mathematical relationship, may be explained thus: if we square the time taken by a planet to return to the same position in its orbit, in other words its revolution period, the result of multiplying the square of the revolution period by the cube of the major axis is a constant, this being true whichever planet is taken and whatever the dimensions of its orbit may be. This last law, then, governs the period of revolution or 'year' according to the planet's distance from the Sun.

The principle of universal attraction, together with Kepler's laws, gives us a complete picture of how the planets move around the Sun, and the satellites around their primaries. But although one body may be gravitationally subordinate to another, it does not necessarily follow that only the attraction of the major body must be taken into account. For example, a planet does not in fact move with the perfect regularity which the above account of its behaviour would lead us to imagine. We have already seen that the shape of the terrestrial globe plays some part in determining the movement of its satellite, the Moon. We must take into account the fact that, although a planet moves generally according to the conditions imposed upon it by its distance from the Sun, it does nevertheless suffer some influence from the other planets, at the same time influencing them in the same way. This causes an alternate acceleration and retardation, and these irregularities, known as perturbations, must be taken into account when calculating the exact position of a heavenly body.

The complexities of celestial mechanics have furrowed the brow of many a mathematician; but later we shall see what remarkable discoveries can be made by analysing such perturbations.

The astronomical distance scale. The distances separating the various heavenly bodies are so great that many people find them difficult to appreciate from the start. Our normal values, based on common terrestrial measurements, prove entirely inadequate when applied on the astronomical scale. We are used to reckoning distances in miles, perhaps hundreds or thousands of miles, and the journey of 25,000 miles round the Earth is considered long. But numbers like these are absolutely insignificant compared with those we have to use in astronomy; and, indeed, 'astronomical' has almost become a synonym for 'enormous'. It is true though that some can be compared with distances found in everyday life: for example, the distance separating a close satellite from its primary. But apart from these we are generally dealing in millions, hundreds of millions, billions and trillions of miles and more, so much so that we must look for some better way of comprehending the real distances between the stars. Nearest is the Moon; then come certain planets,

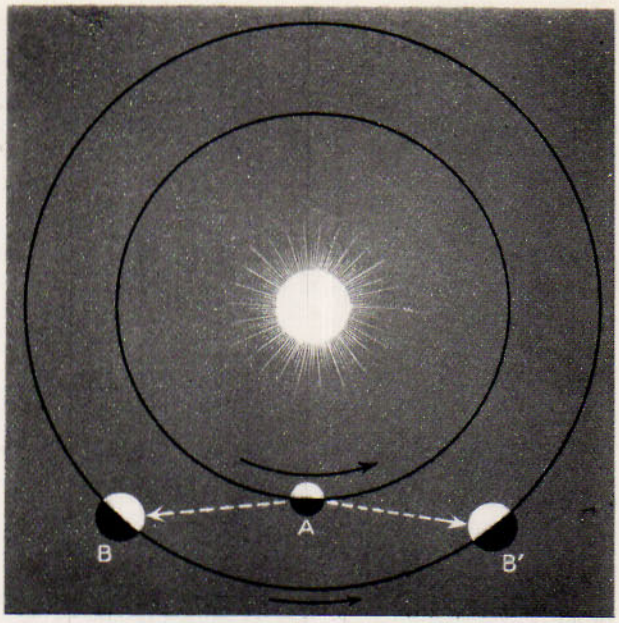


Fig. 32 — PRINCIPLE OF PERTURBATIONS. If two planets are respectively in positions A and B, planet A is slowed down; conversely, when they are at A and B′, A is accelerated.

and the Sun. The Earth is a member of the Sun's family, the Solar System, whose known width is about 6,000 million miles. Though this may seem vast, it is minute compared with the distance to even the nearest star. Therefore, instead of using a meaningless string of noughts, we take advantage of a physical constant to make things clear: the time that light takes to travel a given distance.

An everyday example will be useful here. When a person fires a gun or hits a hard ball with a bat at some distance from where we are standing, there is an appreciable delay between the moment we see him do it and when the sound reaches us. This is because sound takes a finite time to travel a certain distance, its speed being about 700 m. p. h. or 330 metres per second. Many people use this effect to judge how far away a thunderstorm is, by counting the seconds which elapse between the time they see the lightning flash and hear the thunder. We think we see the flash at the exact moment it was produced, but in fact this is merely because the distance the light has to come is extremely small compared with its speed. Light is propagated with a velocity of 300,000 k. p. s. or 186,000 m. p. s. and compared with this 186,000 miles the distances we experience in everyday life are insignificant. Therefore we are accustomed to think that a ray of light emitted nearby travels instantaneously. But the matter is entirely different when we consider the stars. The distance between them and the Earth is so great that a ray of light will take a very considerable time to reach us, as in the case of the sound waves above.

And so we can now bring home the real immensity of the universe. Remaining for a moment within the bounds of the Solar System, we shall begin to realise how large it really is when we learn that a ray of light takes more than 11½ hours to cross it from end to end, whilst it only takes 8m 18s to travel from the Sun to the Earth. May we not truly say that this is nothing in comparison with the gulf separating us from the stars, when we are no longer counting in days, but in years and centuries? Even from the nearest of these other suns a ray of light takes more than 4 years to reach us. There are others from which the journey lasts for 20, 50, 100, 1,000 years and more. And

Fig. 33 — DECREASE OF AN OBJECT'S APPARENT DIAMETER AS A FUNCTION OF ITS DISTANCE

we know stars, at the very limits of our present instruments, whose light must have taken hundreds of millions of years to reach us. Nevertheless we must admit that the thought of the countless aeons during which light has been rushing through space at this incredible velocity may daunt our spirit even more than trying to represent the actual length of the journey by writing it down as a long series of noughts.

In order to simplify the figures, astronomers make use of a unit corresponding with the distance light travels in one year; this unit, the lightyear, is equal to 9,468,000,000,000 kilometers or 5,880,000,000,000 miles. Other units are also employed and will be discussed with the techniques of measuring stellar distances.

The apparent diameters of the heavenly bodies. — Knowing the facts which have been given up to now makes it less surprising that the unaided human eye is not powerful enough to reveal the true nature of the heavenly bodies, despite their enormous size. This will be brought home even more sharply when we study them individually. With the naked eye we can see the Sun and Moon as two large discs, but it requires the magnifying power of the telescope to show the planets as anything more than points of light. The stars are too far away to be directly resolvable even by the most powerful instruments in our possession. However, in all these cases, what we actually see is qualitative in nature: that is we observe them as spherical bodies but their relative sizes as shown by our telescopes do not correspond with their actual relative sizes.

In fact, the discs of the Sun and Moon are, at first sight, about equal in size. We see this beautifully demonstrated at the time of a total eclipse, when the Moon passes directly in front of the Sun. This apparent similarity in size is due to the very different distances of the two bodies from the Earth. We all know an object seems to get smaller as we go away from it. This is an effect of perspective, and it is important not to confuse the object's real diameter with its apparent diameter, which as we have seen, varies as a function of the distance from which it is observed.

Let us recall the 'celestial vault' illusion, whereby all the heavenly bodies appear to be placed side by side in the same plane although they are actually scattered along our line of sight. Now, our neighbour, the Moon, which is relatively small, has the same apparent diameter as the Sun, which is really much larger but also much further away.

It is the same for the various planets. The apparent diameters revealed by the telescope in no way correspond to their real relative diameters. Thus, owing to its proximity, Venus, whose diameter is only one-twelfth that of the much more distant Jupiter, can appear much the

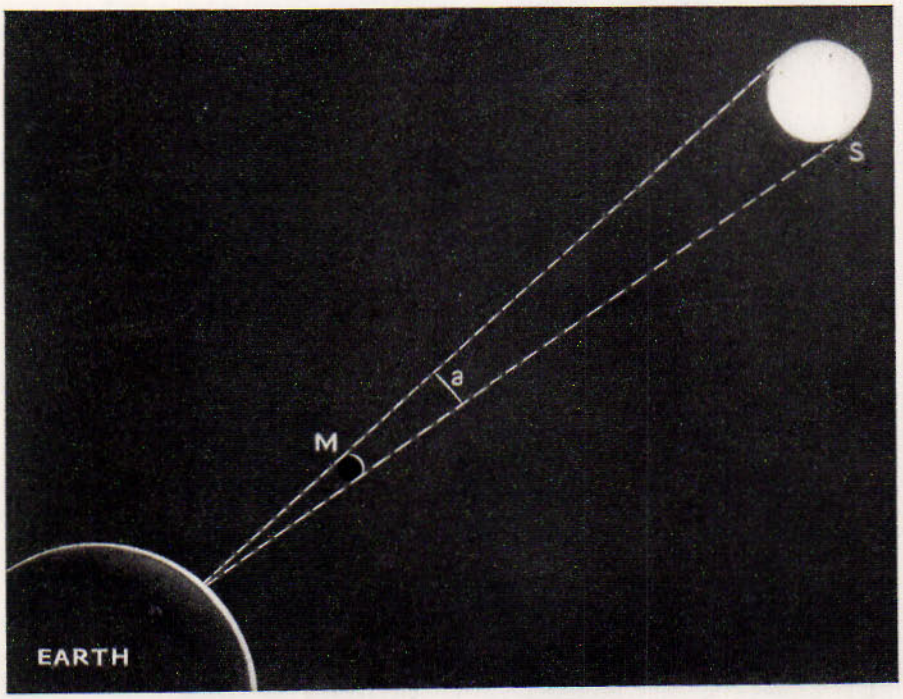


Fig. 34 — THE MOON (M) AND THE SUN (S), unequal in size, have the same apparent angular diameter (*a*) because of their different distances from the observer.

larger of the two. We say 'can appear' because the apparent dimensions undergo considerable variation during the course of the planet's journey round the Sun, which periodically alters the distance separating us.

The apparent dimensions of an object are defined by its angular diameter, that is to say the angle subtended by rays of light coming from opposite sides of the object, and it is expressed, like all angles, in degrees and fractions of a degree. The heavenly bodies are so small, however, that they subtend angles which must be measured not in degrees but in minutes and more often in seconds of arc, shown respectively thus: degree — °, minute —′, second — ″. One degree equals 60′; one minute equals 60″; therefore one degree equals 3,600″. The largest angles are subtended by the Sun and Moon and amount to 31′ approximately. In passing, we may not realise that these often give the impression of being much larger on the horizon than in the sky. If we ask someone to estimate the size of the Moon, for example, when it is just rising, he may reply, using different standards of comparison, that it is about as big as an orange, or a saucer or plate. Allowing for the fact that most people are entirely unaccustomed to making such estimates, it is fair to conclude that to the naked eye these objects appear to be of quite sizable proportions. However, when we go into the matter, it is difficult not to be surprised at how small they really are. A simple experiment will confirm this.

Applying the principles mentioned above, we can take some small object and by holding it comparatively near our eye, make it mask out some very much larger but more distant one. Now what size will just completely mask out this plate-sized Moon that people see? Just a little disc, about one-fifth of an inch across, or a slim pencil of about the same size, held at arm's length (see Fig. 35). If this seems surprising, try it and see. This experiment will also prove that the Sun and Moon are not really any larger on the horizon than high in the sky, in spite of what we think we see. That turns out to be an optical illusion which, although very common, is not as yet fully understood.

Returning now to the appearance of the planets, the largest of them range from 50″ to 60″ in diameter, others are about 18″ to 25″ whilst the smallest are under 4″. Now, to visualise what this means, one second of arc is the angle subtended by one millimetre at a distance of 206 metres from the eye. This simple fact brings home the complete inadequacy of our unaided vision when we try to find out what the planets and stars actually are, and it shows how essential is the help of a powerful telescope if we are to study other worlds fruitfully. The fact that we see worlds thousands of miles in diameter

Fig. 35 — The Moon can be covered with an object about a fifth of an inch across held at arm's length.

simply as tiny discs is yet another indication of how far away they really are. These elementary notions should lead us to a better understanding of their general distribution in space, and the main characteristics of the arrangement which forms the visible universe.

The solar system. — The Solar System comprises all the planets and satellites moving together round the Sun; and the comets may be added to their number, since they behave in the same way.

Let us consider the orbits of the planets as physical things, as race tracks for example. Then from a distance great enough for them to be seen as a whole, the various orbital lanes all appear concentric with one another, all, with a few exceptions, lying more or less in the same plane. But although they are really ellipses, their eccentricity is such that it cannot be shown in the diagram, except in the case of one or two comets. So, for the time being, let us think of the orbits as circles, each with a radius which we shall call the planet's mean distance from the Sun. If then we wish to picture the arrangement of the Solar System in space, we shall have a diagram showing the circular orbits in perspective, rather as a distant spectator on the terraces might see the various lanes of a sporting arena. The planets go round in almost the same plane, whilst the comets for the most part travel in orbits whose inclination from this plane may vary considerably.

The various elements of the Solar System — distances, shapes of orbit, periods of revolution and rotation, etc. — are all discussed in the appropriate chapter. For the moment, let us take a broader look at the size and situation of this system, as it floats in space amidst the stars.

As we have already said, the known diameter of the

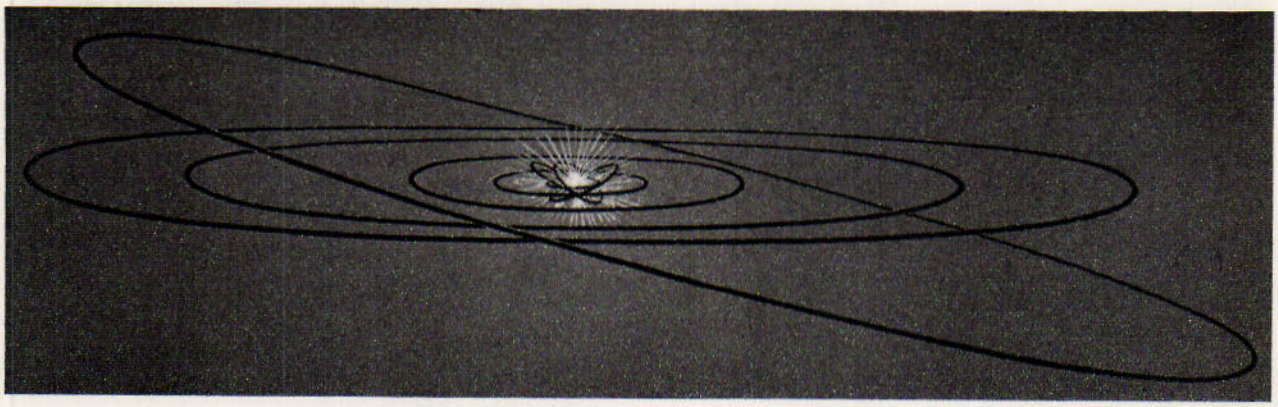


Fig. 36 — INCLINATIONS OF VARIOUS ORBITS IN THE SOLAR SYSTEM

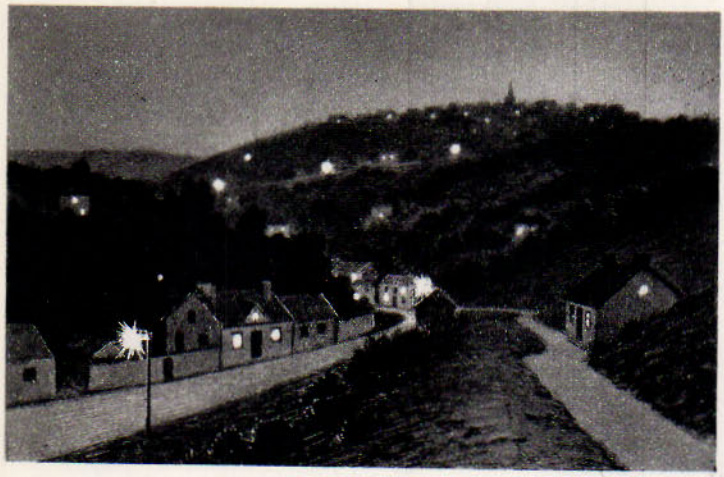


Fig. 37 — View of a town at dusk, showing that the lights are at varying distances.

Fig. 38 — The same lights, seen by themselves, form a 'constellation', with 'stars' of differing brightness.

Solar System, taking the mean diameter of the largest planetary orbit (that of Pluto), is about 6,000 million miles. But this may be extended considerably if we take into account the orbits of various comets, for many of these mysterious visitors — as we shall see later — travel along such highly elongated paths that, quite possibly, many may never have been seen from the Earth (at least since men have been measuring the heavens) or perhaps have made only one return during the whole of recorded history, the size of their orbits never having been accurately determined. Since many cometary orbits are quite inclined to the plane of the Solar System, it follows that we do not know the exact extent of the system either in this plane or in any other direction.

In any case, the boundaries of this collection of planets, however large they may appear, are nothing compared with the immensities of space between the stars.

Let us imagine ourselves making a map of the Solar System and the nearest stars to scale. Even if we used a sheet of paper many hundreds of feet wide, the stars would still have to be marked out round the edges whilst the whole Solar System would be nothing more than a tiny dot lost in the centre of the sheet. That is the scale on which the universe is built.

The stars and the constellations. — Whenever we look at the stars scattered across the breadth of the sky, we are struck by the way some of them, nearer together perhaps than the rest, seem to form definite groups, some irregular in outline, others sometimes having a quite remarkable geometrical regularity. We call these groups the constellations. But this grouping is only an optical effect. It is no use looking for some association comparable with our Solar System, for example. It is true, as we shall learn, that some of these distant stars are not isolated, but are associated in groups sharing to some extent the same destiny, and moving together in the same general direction in space. But the various shapes we call the constellations can be attributed merely to the effects of perspective. Whenever we think we see a group of stars, with certain exceptions we find that they are really scattered through space at very different distances along our line of sight. A familiar example will bring this home. Let us imagine ourselves looking out from a moderate distance over a town with all its street lamps lit up in the twilight. It will of course be quite easy to see the relationship between all the points of light and their respective distances by means of various landmarks and lines of convergence.

Now suppose that night falls, leaving us unable to see anything but the various points of lamplight. (A snapshot would be better for this purpose because it would not pick up all the patches of diffused light that our eyes would make out.) All impression of depth would be lost owing to the disappearance of our reference points and we should be left looking at a series of lights, of different intensities, all apparently in the same plane, forming clusters here and there purely by chance.

And so the constellations are simply this random clustering of stars which, although looking close together, may be hundreds of light-years apart.

In ancient times these constellations were given symbolic or mythological names which have persisted down to the present day. Whilst quite arbitrary, these names nevertheless have the advantage of making the sky much easier to describe by cutting it up into convenient divisions. The stars of each constellation are designated in order of descreasing magnitude by a Greek letter: Alpha (α) followed by the genitive of the Latin name of the constellation, is the brightest, Beta (β) the next, and so on. But most of the stars, at any rate the more brilliant ones, were also given names of their own by the ancients, and these proper names are still in use. So Sirius, the brightest star in the sky, is also Alpha Canis Majoris; and Vega is Alpha Lyrae, etc. Another more scientific nomenclature has been added to these earlier but still useful ones and we now catalogue the stars according to their apparent positions in the sky. These positions are measured with respect to an imaginary system of longitude and latitude established on the celestial sphere, much in the way we measure the position of points on the Earth's surface. Many stars, especially the fainter ones, are identified solely by the numbers assigned to them in the different catalogues.

The visible universe and our position within it. — Let us now review what has been said in the preceding sections to give us a broad general survey of the universe before going on to examine each part of it in detail. First, we live on a planet which circles its primary, the Sun, in company with eight others, many attended by a retinue of satellites. This system is large, yet small. At a tremendous distance beyond its furthest limits lie the stars, those other suns in the depths of space. Effects of perspective form these into groups — some, the constellations, being loose and indefinite, others, more compact making up the clusters and bright clouds of the Milky Way.

If it were only a matter of their respective distances, there would seem to be no relation between the stars at all. But this is not so. All the stars we see, numbering several thousand million, belong to one great system floating in space. For, although telescopic observation reveals many more stars beyond those we can see with the naked eye, this does not mean that they go on indefinitely. Then we can consider the limits beyond which there are no more stars; but for the moment we shall leave this subject, one of the vastest we can conceive, and that we ourselves are plunged in the thick of the crowd. We are not in the middle but occupy a somewhat eccentric position near the outer rim; this explains why we see a greater or lesser accumulation of stars according to the direction in which we look out: remember what was said about the cause of the Milky Way.

Remember, too, that the Earth is a member of the Solar System and that the Solar System is totally insignificant compared with the size of the Galaxy: so when we say 'the Earth', it is just as though we were to say the

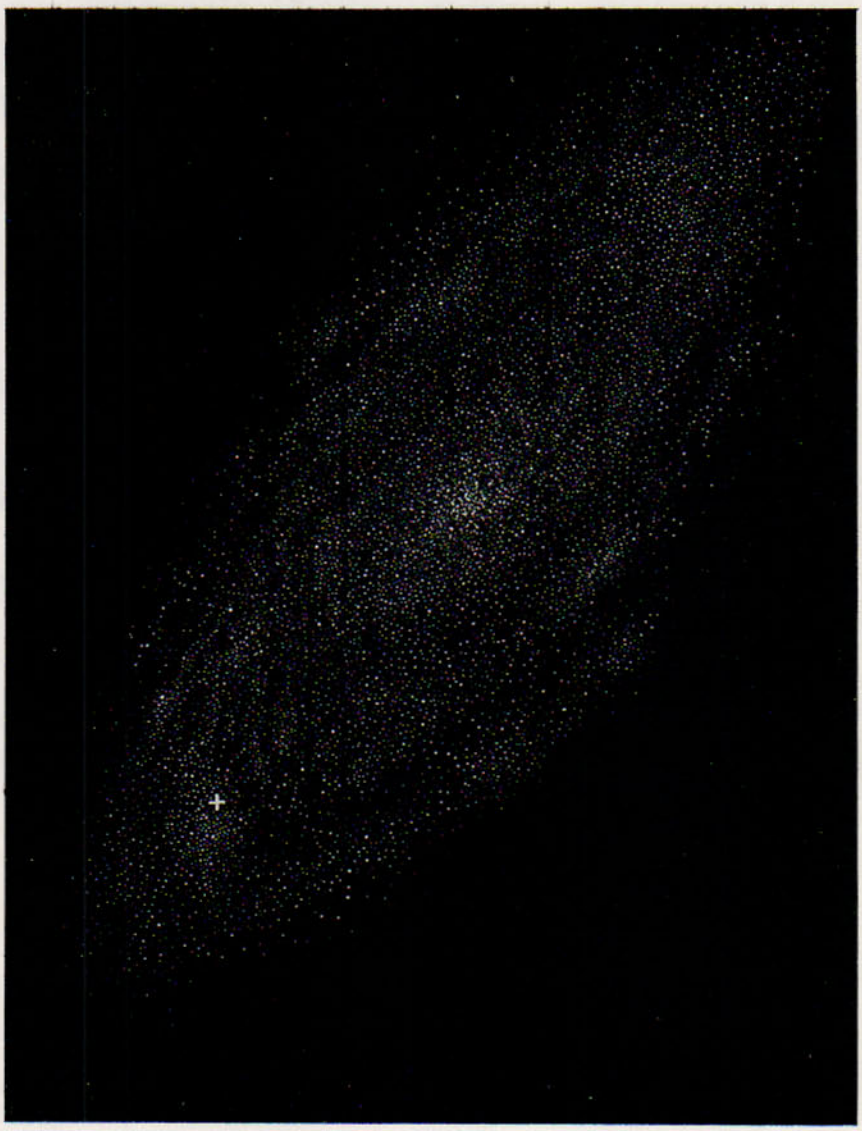


Fig. 39 — OUR STELLAR SYSTEM, THE GALAXY, IS PROBABLY OF A FORM RESEMBLING THIS. The position of the Solar System is indicated by a cross.

be content to sketch out the form of our own stellar system.

The immense accumulation of stars and gas we call the Galaxy is roughly lenticular in shape. The material out of which it is made is not equally distributed everywhere but is condensed into knots of various sizes. We may think of the Galaxy, with its flat, circular form and its slow rotation about an axis perpendicular to the plane of the 'lens', as a sort of enormously enlarged version of the Solar System.

Since we see individual stars all around us, this means point in space occupied by the Sun, so far as the rest of our stellar system is concerned. The Sun at the heart of this mass is simply a member of the Galaxy, a star just like millions of others, as unremarkable as an individual lost in the midst of a huge crowd. For the Sun is neither the largest, nor the smallest, but just about average, and the position it occupies is absolutely unprivileged. Sharing the same fate as its countless companions, the Sun makes its journey at high speed, accompanied by its retinue of nine planets, their satellites, the minor planets, the comets, meteors and gas, as well as all the other

cosmic debris which goes to make up the Solar System.

Some of these galaxies we have mentioned are so far away that their light takes millions of years, even as much as 2,000 million years, to reach us, whilst it would take only about 100,000 to cross our own Galaxy from end to end. So compared with the distances separating them, the diameters of these galaxies (or spiral nebulae) are minute.

That, broadly speaking, is the known arrangement of the universe we see as the 'sky'. We know that everything is in movement (by virtue of which everything changes in its season) and we know what we can expect to see. Now we are going to find out just how some of these sights will change, and how others will recur periodically according to how we examine them from our observatory, the Earth.

Fig. 40 — OTHER GALAXIES RECEDING INTO THE DEPTHS OF SPACE

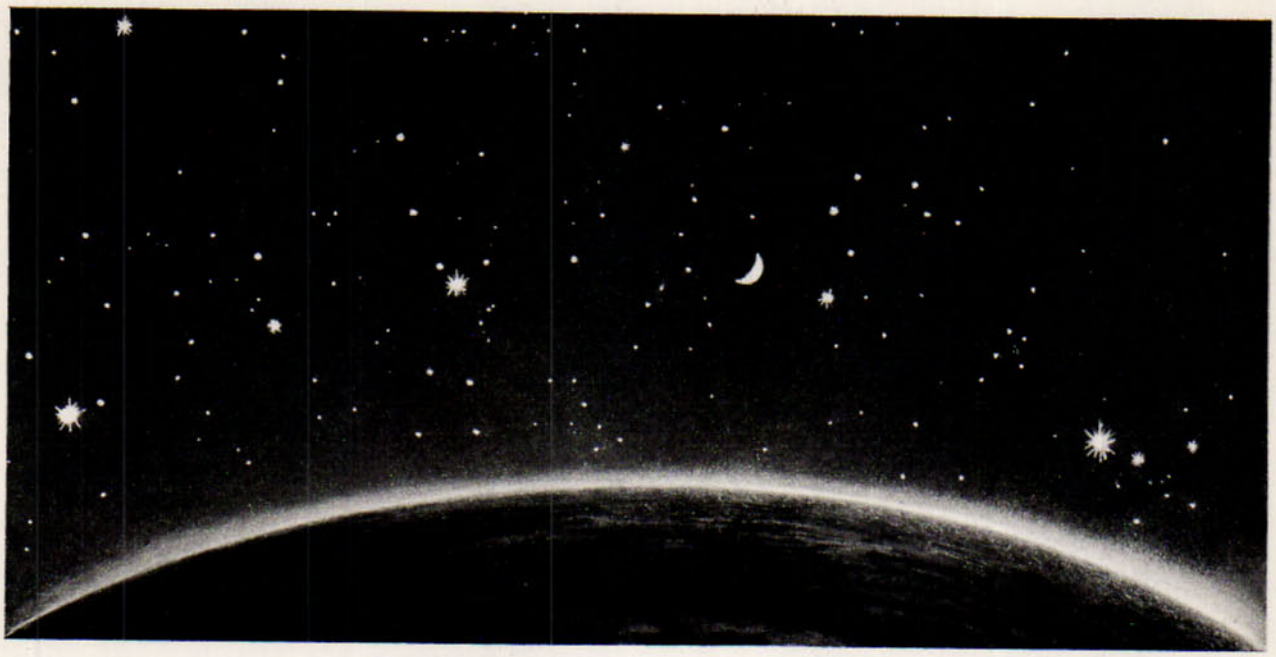


Fig. 41 — STARS AND PLANETS AROUND THE EARTH

CHAPTER 2

OUR OBSERVATORY, THE EARTH

The rôle of the atmosphere. — In the first chapter we looked at the Earth, Moon and planets from various points of view, but mainly as bodies moving through space in ways determined by the great universal laws. It is by emphasising the planetary nature of our world that we shall best understand the successive and periodic phenomena that its movements give rise to: for example the cycles of night and day, winter and summer. And at the time when the expansion of human knowledge began, it was these sometimes deceptive appearances which gave experimental proof that the Earth is round and does not need to be supported in space as the ancients believed.

Details of the Earth's dimensions, its exact shape, the various materials of which it is composed, its interior and surface conditions, are the subject of a special chapter (Chapter 4), so for the moment let us limit our account to the effects resulting from the conditions under which we survey the surrounding universe of stars.

From the strictly cosmic point of view, it does not very much matter whether the Earth has an atmosphere or not, but it matters a great deal to us! This envelope of air, outside which we could not survive without taking extremely elaborate scientific precautions, plays an important part in determining the conditions of our daily life.

The fairly thick layer of gas we call our atmosphere forms a medium interposed between us and the stars, and it inevitably influences our way of finding out what lies beyond. To start with, the atmosphere affects the appearance of the stars and planets: for example their light may be considerably dimmed and is frequently cut off altogether by obstacles in its path; and an astronomical image suffers refraction in the atmosphere, which has the effect of displacing the object's apparent position and upsetting the precision of our attempts at measurement.

Although we may blame the atmosphere for a great deal, we should be grateful for the effect it has on the Sun's light. Were it not for this protection, the Sun would shine on us with a fierce intensity that would have many unpleasant consequences: for example, human tissue would be damaged by the strong ultraviolet radiation which is now cut off by the ozone above our heads. Almost blinding in its brilliance, the Sun would beat down upon us from a pitch-black sky where we should be able to pick out the stars in daytime if we shaded our eyes, and this would serve to emphasise the fact that the Earth is indeed a body travelling through space. In short, we should feel as if we were being grilled to death by the fierce radiation of an arc-lamp with nothing to moderate its harshness.

The Earth's atmospheric covering partly diffuses the sunlight before it can reach the ground, thus turning itself into a brilliantly lit veil; although preventing us from seeing out into space, this veil of light bathes us in the general illumination we call 'daylight', penetrating even the thickest clouds which hide the Sun. We can demonstrate these effects in the following way: place a bright, unshaded lamp on the sill of an open window at night. You will see it as a brilliant source of light against a black background. Then stretch a fine gauze curtain in front of the window between you and the lamp: you will be able to see the lamp through the weave of the material, but the dark rectangle of night will be lost behind the light diffused in the curtain, and this light will be spread throughout the room as a general glow

Fig. 42 — THE ATMOSPHERE GIVES AN EFFECT OF PERSPECTIVE TO A LANDSCAPE

Fig. 43 — THE SAME LANDSCAPE WITHOUT ATMOSPHERE

much more gentle to the eye than the naked lamp. Then put out the lamp and you will be able to see the lights in the distance through the transparent curtain. In principle this is the way the Earth's atmosphere behaves.

The atmosphere is remarkable, too, for the fact that sunset is not immediately followed by the blackness of night: and, by the same token, sunrise is heralded by an upward-spreading radiance in the morning sky. This is

Fig. 44 — WE ARE CUT OFF FROM THE STARS BY THE EARTH'S ATMOSPHERE

the phenomenon known as twilight. Its cause is simple and easily seen from the diagram, which shows that even though the spot where we are standing may receive no direct light from the Sun, it receives an indirect illumination from the appreciable thickness of atmosphere overhead still catching the solar rays. The twilight glow draws steadily down towards the horizon, disappearing when the Sun has reached an angle of 18° below. This will be dealt with in more detail later when we consider the double effect of the Earth's rotation and the seasons.

The earth in space. — The best demonstration of the Earth's isolation in space is that we can see the stars climb up over the horizon (which delimits the part of the globe we can see from any one place), travel across the sky, and disappear on the opposite side. The next day they reappear at the same spot, repeating the same cycle indefinitely. This is the effect of watching stars which are going round the sky in circles from the surface of a great opaque body which hides from view all the circles except the part in front of us. We are rather like a man in a stand built in the centre of a sports arena, able to look out in one direction only. All but a part of the track is out of his sight so that, if he is watching a race, he sees the runners only as they pass across his short section of track again and again. If our spectator now takes a walk round his stand, the rest of the track will begin to come gradually into his view and he will finally realise that it forms an uninterrupted circuit. This is exactly our situation on the Earth: wherever we go, from whichever part of the Earth we see them, the stars will seem to go round the Earth without a break in their path.

Let us now examine this circular movement in its entirety. Instead of looking out in any chance direction, let us direct our gaze directly north or south according to which hemisphere we are in. We now find that certain stars no longer disappear beneath the horizon: they curve down towards it, coming very near at the lowest point of their course, then sweep up again with majestic slowness. There is no pivot in the sky round which the stars circle: they turn about an imaginary centre — the celestial pole — in circles whose radii become larger the further they are from the pole, rather in the manner of a compass being opened at successively wider angles. A photograph of these curves gives a very vivid demonstration of the circling heavens. A star at the centre of rotation would appear stationary, and in the northern hemisphere we can indeed see such a one occupying the privileged position. Being of moderate brightness the Pole Star (Polaris) serves as a useful direction finder for someone without a compass, but it has more than a practical value. Although is seems to remain in the same place, this is simply because it is so near the pole; in fact it describes a complete circle round the true pole during the course of a day, but this circle is so small that the movement is hardly noticed by the naked eye.

All this indicates a freely moving Earth, unsupported in space by the natural pivot which the ancients believed to exist.

The idea of the Earth as a sphere isolated in space is

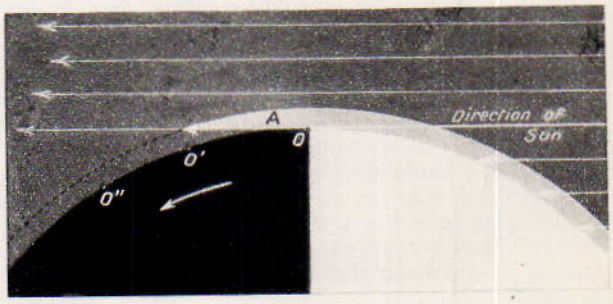


Fig. 45 — FORMATION OF TWILIGHT

Fig. 46 — APPARENT ROTATION OF THE SKY AROUND THE CELESTIAL POLE

demonstrated physically by the beautiful phenomenon of a total eclipse of the Moon, which directly shows the shape of our globe. As we have seen, an eclipse of the Moon occurs when it passes behind the Earth with respect to the Sun; in other words, by moving into our planet's shadow, the Moon cuts herself off from the source of light by which she shines and therefore disappears from sight. As the edge of the Earth's shadow passes across the Moon's disc, we can see that it forms part of a large circle — the true shape of the Earth.

After describing these apparent movements, we must now see what really happens. The rôles must be reversed: it is not the sky that rotates freely about a stationaryEarth, as was believed for many centuries: it is that the Earth turns on its own axis.

The diurnal movement of the stars is therefore just an illusion, the counterpart and, in a sense, the opposite of the real movement with which our world is endowed.

Fig. 47 — THE SHADOW OF THE EARTH ON THE MOON'S DISC (eclipse)

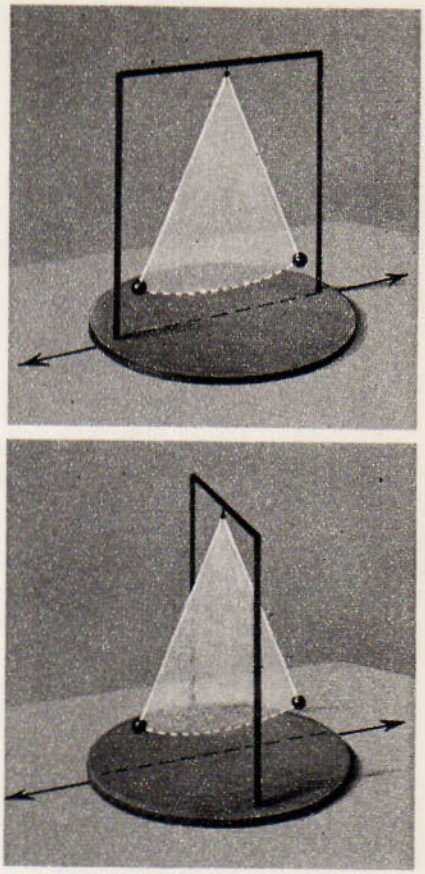


Fig. 48 — PRINCIPLE OF FOUCAULT'S PENDULUM

Earth's rotation, which we do not feel because our bodies are swept along at the same time as the ground on which they are standing, can be proved by means of a celebrated experiment devised by the French physicist Léon Foucault. One might call this a tangible proof.

Foucault's pendulum. — A remarkable property of the pendulum is the invariability of the plane in which it oscillates. This can be shown with an extremely simple piece of apparatus.

A small pendulum, consisting of a thread with a ball of lead attached to one end, is freely suspended from a horizontal bar mounted on a base which can be turned in the horizontal direction. If, with the apparatus in one position, we start the pendulum swinging, it will oscillate in a plane defined by the two extremes of its swing (Fig. 48). Then, if we gently begin to turn the base of the apparatus, we shall find that this movement does not affect the direction of the pendulum's swing: in other words, the plane of oscillation will always point in the same direction. This is the principle applied by Foucault to demonstrate the Earth's rotation. His reasoning was as follows.

Since the plane of oscillation of a pendulum constitutes a fixed standard of reference for the movement of surrounding objects, a pendulum suspended from some support fixed to the Earth must appear to change its direction of swing with respect to the support if the latter is being carried round in space by the movement of our globe. In 1851 Foucault undertook such an experiment in the cellar of his Paris house, and the success he achieved caused a widespread sensation. In the following year therefore, this experiment was performed again, with more pomp and ceremony, in the Panthéon, in Paris (Fig. 50).

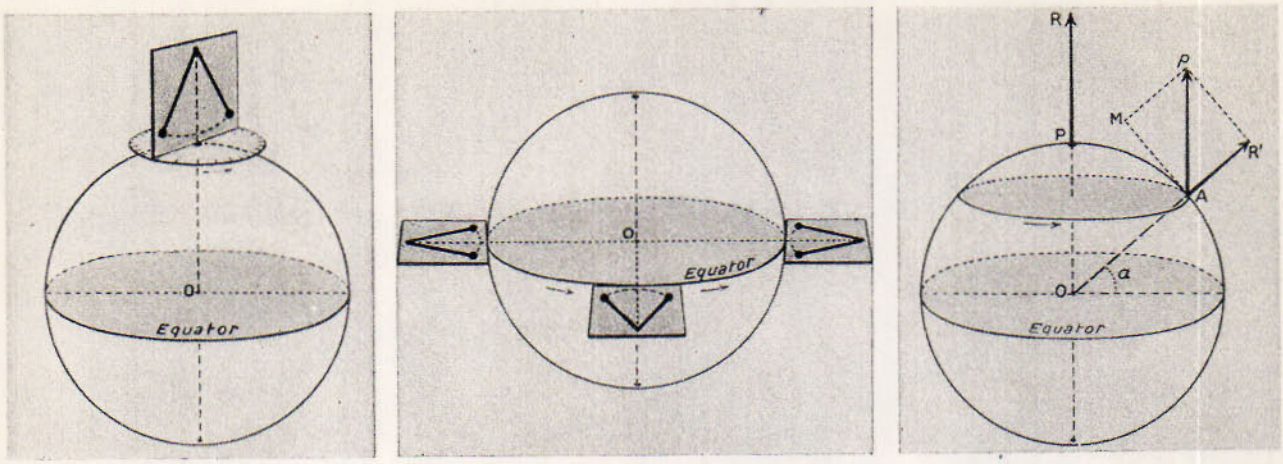


Fig. 49 — FOUCAULT'S PENDULUM EXPERIMENT
(a) At the pole (b) At the equator (c) At an intermediate latitude

This time, the pendulum consisted of a bronze sphere weighing about 56 pounds, with a spike on the underside. It was suspended by a steel wire from the topmost point of the dome. At the end of each swing, it was arranged that the point of the pendulum should touch the top of a heap of sand, and make in it a furrow exactly defining the direction of swing. As this direction remains invariable in space, the Earth's rotation displaced the sand beneath this fixed standard of reference, and since the displacement is appreciable during the time of one oscillation, at each swing the watchers saw a new furrow ploughed beside the previous one. For a spectator behind the pile of sand, the displacement of the streaks was from left to right, which indicated that the sand and therefore the Earth must be moving from right to left. This we know to be the real sense of its rotation.

This remarkable demonstration was repeated on several occasions, notably in 1902 at the Panthéon to commemorate the fiftieth anniversary of the original demonstration. It was organised by Professor A. Berget under the auspices of the Société Astronomique de France, of which Henri Poincaré was then president and Camille Flammarion secretary. There is also a Foucault pendulum at the Science Museum in Kensington, which was installed in its present position about thirty years ago.

Remarkable and instructive as this experiment may be, it must be pointed out that the information to be gained from it is purely qualitative in nature; this means that, although it vividly demonstrates the Earth's rotation, measurement of the displacement shown by Foucault's original pendulum in Paris would lead us to conclude that the Earth's period of rotation was about 32 hours instead of 24, which is the correct period determined by means we shall discuss below, while the pendulum at the Science Museum in London could give an apparent rotation period of slightly less than 31 hours.

Why the discrepancy?

It is because at the latitudes on which London and Paris lie, the pendulum's plane of oscillation is perpendicular to the Earth's surface at that point and does not coincide with the vertical represented by our planet's axis of rotation. It is in this difference that the cause of the observed discrepancy lies.

If we performed the experiment at the north pole, the vertical of the pendulum's suspension point would coincide with the rotational axis of the globe and the Earth's surface would therefore make one complete turn, measured with respect to the pendulum's plane of oscillation, in 24 hours, or conversely, the latter would make a one complete turn in relation to the ground in this same time. If, on the other hand, the experiment were made at the equator, the vertical would be perpendicular to the axis of rotation. In order to make the accompanying diagram clearer we have shown the plane of oscillation as a small square orientated in the plane of the equator, which is always perpendicular to the axis of rotation. Thus the pendulum would show no signs of deviating from its path, and the point would keep passing along the furrow ploughed in the sand at its first swing.

Between the apparent speed of rotation equal to that of the Earth at the poles, and a zero apparent speed at the equator, we should expect to find at an intermediate latitude an equally intermediate rotation period: that is a speed somewhere between the actual speed of the Earth as measured at the poles, and a zero speed, meaning an infinite rotation period, at the equator.

What really happens? The principles of mechanics show that rotations are composed, like forces, according to the so-called 'law of parallelograms'. A rotation is represented by an arrow in the direction of rotation, the length of the arrow being proportional to the speed of the movement. Applying this principle to the Earth, the speed of rotation will be represented in Fig. 49*c* by the arrow PR. At point A, of latitude *a*, whose vertical is the projection of OA, we want to know how the pendulum will show the Earth's rotation not about the vertical to the axis of rotation, but about OA. From point A, draw an arrow parallel with and equal in length to the one showing the rotation, and another parallel to the plane of the surface. Now draw a parallelogram with diagonal A*p*: the length of AR, corresponding to that of the smaller side defined by the vertical at A, gives, in proportion to the length of A*p*, representing the real rotation period, the observed period at A.

Mathematically, this explanation would be expressed as follows: the angular speed of rotation observed and measured by the displacement of the pendulum is equal to the real rotation speed of the Earth multiplied by the sine of the latitude. And, in fact, when we calculate this for the latitude of Paris, we find that the period should be 32 hours, which is in exact agreement with Foucault's observation.

The Earth's rotation period. — We have so far been discussing the effects of the Earth's diurnal rotation. Now we must define its duration with greater precision, for this must not be confused with what we commonly call the 'day'.

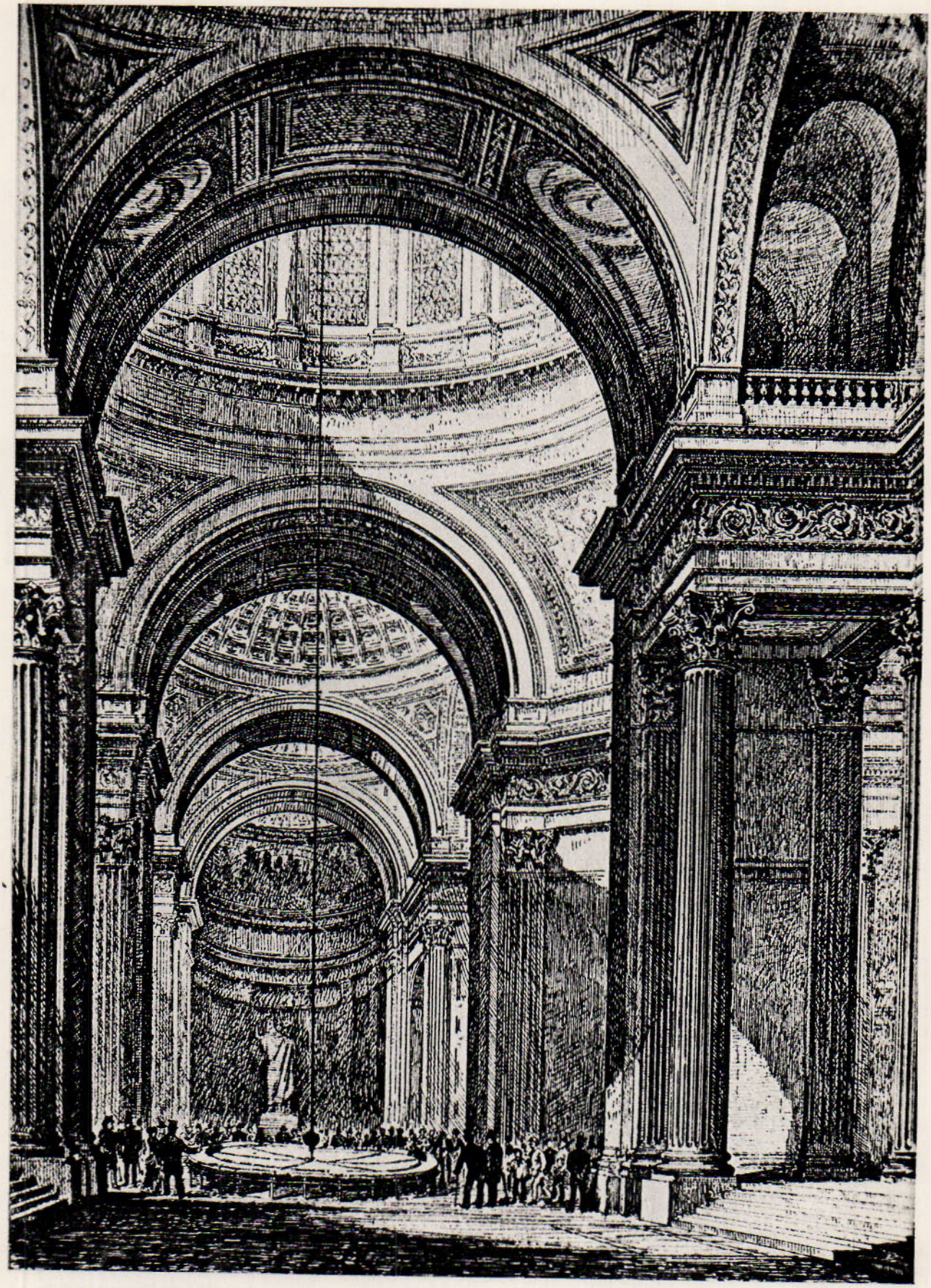


Fig. 50 — FOUCAULT'S DEMONSTRATION IN THE PANTHEON, PARIS. (Courtesy, Science Museum, London.)

For normal purposes a 'day' is equal to twenty-four hours, this being the formally recognised way of dividing time. Below we shall learn that this period is used although it does not represent the real rotation of the Earth as measured on the celestial sphere.

To start with, note the time taken by a star to come back to a given point in the sky. By a perfectly regulated watch it would pass this reference point at intervals of 23h 56m 4.091s. This is the real, or sidereal period of the Earth's rotation.

It will be seen that this period, sometimes called the sidereal day, is 3m 55.909s shorter than the civil day of 24 hours, by which the life of Man is regulated. The cause of this difference will tell us much about the many factors which contribute to the ever-changing appearance of the sky.

The Earth is the chronometer by which the astronomer determines fundamental celestial positions; it is therefore essential to know up to what point he can trust its regularity.

Obviously, when we rely on observation, its working appears perfectly regular, because the stars rise, transit the meridian and set with perfect punctuality. Yet this is not quite true.

It would take too long to recount exactly how we came to doubt the marvellous chronometer which keeps our universal time. By rights, this subject belongs to the specialists in celestial mechanics, and we can merely say that these specialists became suspicious when analysing the movements of the Moon; it was noticed that over a period of time our satellite appears to accelerate, getting further and further ahead of its predicted position. This was called the Moon's secular acceleration.

The explanation of this fact, which puzzled astronomers for a long time, was found by Delaunay in 1850. It is not the Moon which is accelerating but the Earth which is slowing down, the Moon's speed being measured with reference to the Earth's rotation. A more searching and refined analysis has since confirmed this prediction. In addition, Brown and de Sitter have analysed the movements of the planets and reached the same conclusions.

The result of all this detailed research has been to show that the Earth's rotation is not only slowing down but is also subject to smaller irregularities.

Generally speaking, the main retardation can be attributed to a continuous braking action due to friction developed between the ocean masses and the surface of the globe under the tidal pull of the Moon. But this is countered, one might say, by forces inherent in our world itself: changes in temperature, displacements of material, vertical movements of the crust, etc.; these cause slight modifications in its diameter, all of which engender sudden, probably periodic variations in the opposite direction. We can picture the slowing down of Earth's rotation as the temperature chart of a fever, the general curve descending but fluctuating slightly on either side of a mean line.

The results of inquiry into this problem are as follows: if we compare the Earth to a clock, this clock will slow down

By 17.51s	after 100 years;
By 1m 6s	after 200 years;
By 1h 50m 32s	after 2,000 years;
By 1 day	after 7,600 years.

Over long periods, these large perturbations can be treated as if they were of uniform magnitude, the Earth either gaining or losing in turn. In modern times three such periods can be recognised.

From 1660 to 1787 the retardation increased in proportion to the time, reaching a total of 19.2s or 0.139s per year. During this period, the rotation could be treated as remaining constantly 0.00038s less than its mean value.

From 1788 to 1898 there was a gain totalling 18.1s, corresponding to a rotation period increased by 0.00045s per day.

From 1898 onwards, there has been a new annual retardation of 0.153s, the rotation period shortening by 0.00042s per day.

But within the last decade or so, we have detected other, more rapid fluctuations superimposed on the longer variations. Our modern quartz-crystal and caesium clocks keep such perfect time that they can be used to time astronomical phenomena, and this includes the detection of minute irregularities in the Earth's rotation. Analysis of such data by Stoyko at Paris Observatory has revealed accelerations and retardations over the very short period of about one year.

The last word has not yet been said on this matter and there is still much to be done. However, this is a question which need not trouble most of us for it does not affect the setting of our watches. Astronomers, though, justifiably worry about it.

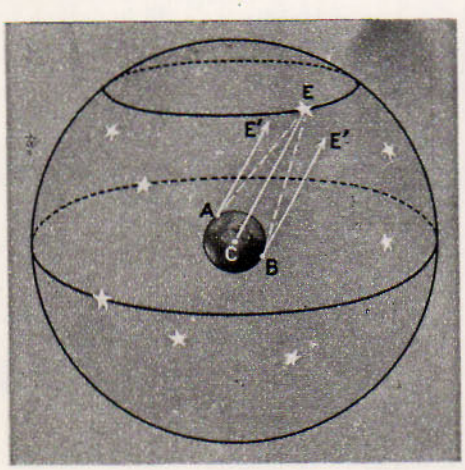


Fig. 51 — POSITION OF A STAR ON THE CELESTIAL SPHERE. E, position of a star as seen from the centre of the Earth, C. From A or B, its direction is parallel, E'.

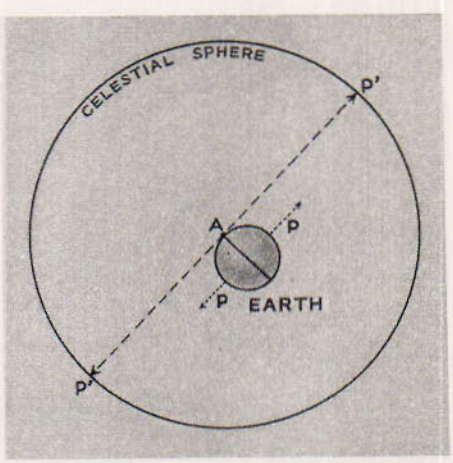


Fig. 52 — THE CENTRE OF THE CELESTIAL SPHERE always appears to coincide with the observer's position, A, on the Earth. PP, Earth's axis; P'P', line of the poles.

Visibility of the sky from the Earth's surface. — Before going on to a further consideration of the diurnal movement of the sky, we must understand how the sky is observed or rather what part of it is observed. From here on, Earth must be treated as a sphere. This in itself enables us to understand things more clearly.

Our point of departure is this simple principle — which should not be surprising after all we have said about the immensity of space outside the Earth — that the size of our planet must be considered as absolutely insignificant compared with the distance to the stars. If we assume that the centre of the Earth coincides with the centre of the celestial sphere, we shall arrive at the same conclusion about the visibility of this sphere from various parts of the Earth's surface.

However, this is not the case with neighbouring bodies such as the Moon, whose distance of only a few thousand miles introduces changes in the respective situations of the points seen (as we observe when, while walking, the details of the surrounding landscape seem to move relative to us). By the application of very simple geometrical principles these effects can be used, for example, to determine the Moon's distance.

But let us return to the general aspect of the imaginary celestial sphere on which the distant stars seem to be situated. In the accompanying diagram (Fig. 53) we may place the centre of this apparent sphere at the point occupied by the observer, whatever position he may occupy on the surface of the globe. But in order to take account of the consequences of this with regard to the sky, we must show the Earth as having an appreciable diameter, and it will thus occupy an eccentric position inside the sphere. Once this idea is accepted, we can show exactly how we observe the sky, believing ourselves at the centre of the celestial sphere, no matter where we really are.

At any one moment we can see only part of the sky: that which is above the horizon.

As an elementary and provisional definition of the horizon, let us take it as the limit of the horizontal plane apparently formed by the surface of the ground on which we stand. Beneath and beyond this limit we can of course see nothing. We could compare it to the floor of a domed building, or better still, a floor bisecting a sphere of which we can only see the top half. In Fig. 54 the plane of the horizon considered thus is represented by a tangential plane at our latitude on the globe; in relation to this, the zenith is the point above our heads which determines the vertical, i. e. the projection on to the celestial sphere of the particular Earth-radius which emerges beneath our feet. In other words, if we divide the celestial sphere into latitude circles parallel with the equator (as the Earth itself is divided), the plane of the parallel on which the observer is placed, projected into space, determines the celestial latitude circle which passes through his zenith.

For any place on the Earth's surface, we can immediately establish the location of the celestial equator and one or both of the poles — which are separated by an angle of 90° — with respect to the horizon. If we suppose that the chosen spot is at latitude 50° north, the zenith will be found at the same degree of celestial latitude, and the pole will be at an angle equal to the difference between this figure and 90°, i. e. 40°. Likewise, the zenith is 90° from the horizon so that the pole must make an angle of 50° with the latter, the complement of the angle it makes with the zenith (Fig. 54).

For all practical purposes, the situation can be summed up as follows: at any given place, the angle between the celestial pole and the horizon is equal to the geographical latitude of the place.

If we change our position on the Earth's surface, that is to say, move to where another terrestrial radius emerges from the ground, the zenith will move by an equal amount on the celestial sphere, the latter then being cut at a different angle by the horizon. To continue the comparison used earlier, it would be as if the floor of our domed building had tilted but we had managed to remain perpendicular to it. We no longer see the same part of the celestial sphere as before; part of it will have disappeared where the floor was tilted upwards, and another will have come into view on the opposite side. Fig. 55 makes this clear.

Now we must consider the behaviour of the stars as determined by their orientation to the horizon, which bisects the celestial sphere. The diagrams show that, although we only see one hemisphere at a time, this need not necessarily correspond to half of the sky. In explanations of this sort, it is common practice to employ geographical terminology; thus the northern and the southern hemisphere are the two halves of the celestial sphere of which the north and south poles respectively appear to be the centres of rotation. The dividing line between the two hemispheres is the celestial equator, which is the

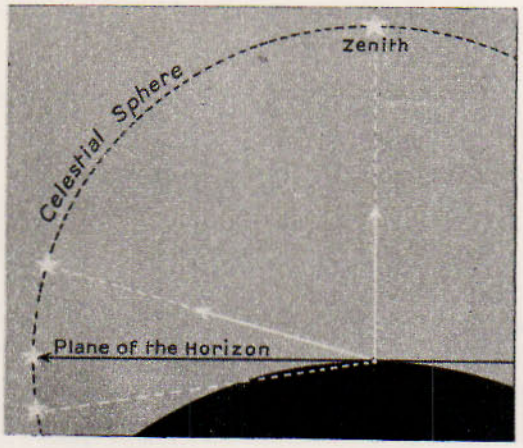


Fig. 53 — PLANE OF THE HORIZON AND THE ZENITH

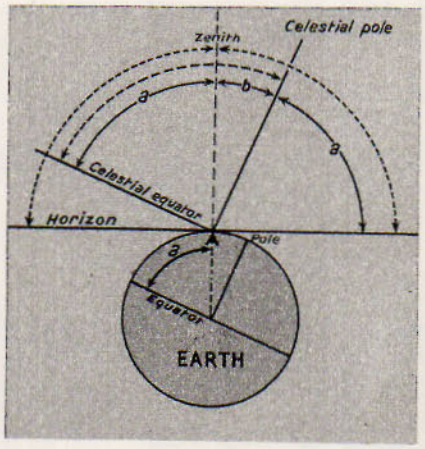
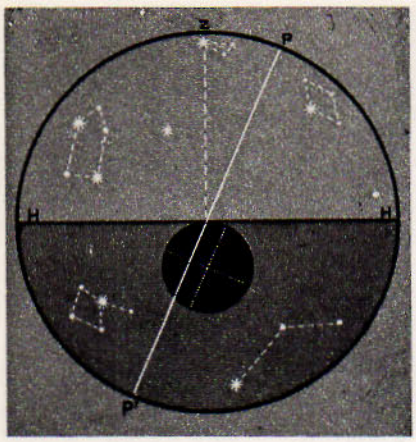


Fig. 54 — CELESTIAL AND GEOGRAPHICAL LATITUDES. The angle (*a*) of the pole above the horizon is equal to the latitude of the place (A).



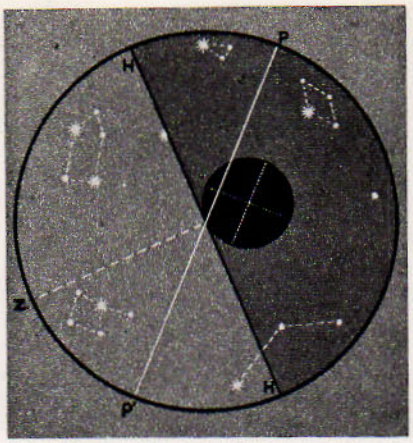


Fig. 55 — THE SKY AS SEEN FROM DIFFERENT POINTS ON THE EARTH'S SURFACE

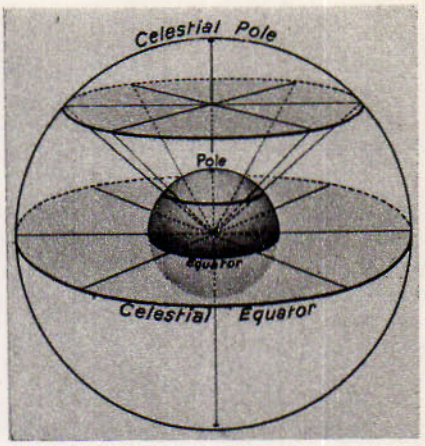


Fig. 56 — CELESTIAL LATITUDES CORRESPOND WITH GEOGRAPHICAL LATITUDES

projection of the Earth's equator on to the celestial sphere.

Furthermore, when we talk about the point in space occupied by the observer, it should be taken to mean the latitude of this place on the Earth's globe. Latitude circles are drawn parallel with the equator, which is the plane of the rotatory movement: thus for all points on the Earth's surface with the same latitude, the apparent movement of the sky will take place in the same way.

Let us see how this movement does appear. First let us imagine an observer exactly at the north pole. Here the line joining the poles would coincide with the vertical, the north pole being exactly above his head; the plane of the horizon, which is exactly perpendicular to this direction, would coincide with the celestial equator. So our observer will see all the stars of the northern hemisphere moving parallel with the horizon, beneath which all the stars of the southern hemisphere will alway remain invisible. The situation is exactly reversed at the south pole. We may recall that in all these cases, the radius of the circle described by a star depends on its distance from the pole, being smallest when closest to the pole.

Next suppose that our observer moves to a spot on the equator, his new position being exactly at right angles to his former one. The line joining the poles will now lie in the plane of the horizon, and he will see the two centres of rotation on opposite sides of him, with the result that at some time or other during the 24 hours he will find, in theory, that all the stars in the sky will come into his view, any and one star will be above the horizon for 12 hours out of the 24.

Lastly, let us take the case of an observer at some intermediate latitude between equator and pole. Now the plane of the horizon is oblique in relation to the celestial equator, and the latter cuts the horizon at two points at opposite ends of a diameter, the stars at these points being visible through one half of their diurnal journey. But this will not apply for the others. Take those near the north pole, which will be at a fairly moderate altitude above the northern horizon; these stars will describe circles of a smaller radius than those at the equator, and the circles will be cut into two very unequal halves by the plane of the horizon. The upper half, which is the star's path above the horizon, will be larger than the lower, which corresponds to the time it is invisible below the horizon.

The nearer the stars are to the pole, the smaller the portion of their diurnal circle which is cut off and hidden by the horizon, until a time is reached when the horizon just makes a tangent with the lowest point on the circle. From then on the stars no longer set, but remain visible all the time and are said to be circumpolar. From what has already been said, it will be understood that the circumpolar zone varies in width according to the latitude of the observer, so that we finally return to the first example of an observer at the pole itself, who has the whole northern hemisphere of the sky as his circumpolar zone.

Conversely, the stars in the southern hemisphere, as

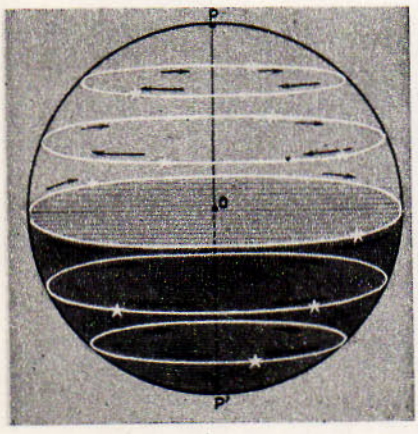


Fig. 57 — APPARENT MOVEMENT OF THE SKY AT THE POLE

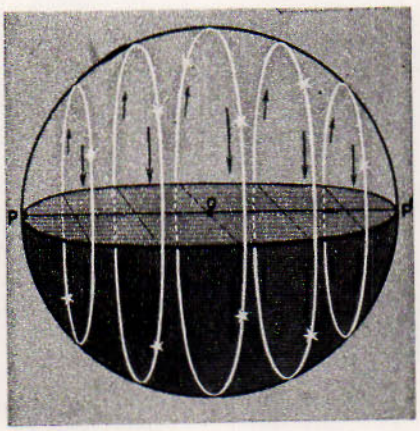


Fig. 58 — APPARENT MOVEMENT OF THE SKY AT THE EQUATOR

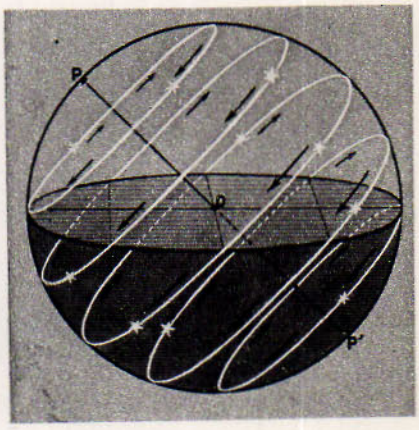


Fig. 59 — APPARENT MOVEMENT OF THE SKY AT LATITUDE 45°

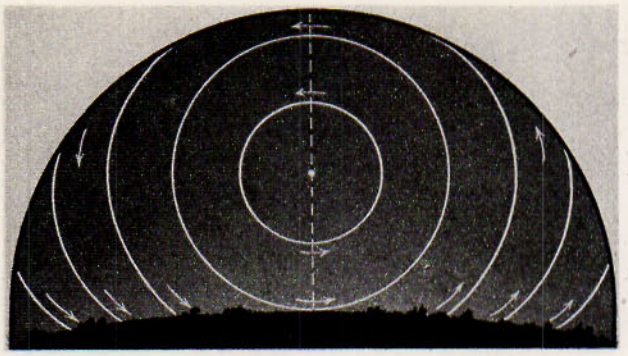


Fig. 60 — APPARENT MOVEMENT OF THE SKY: LOOKING NORTH

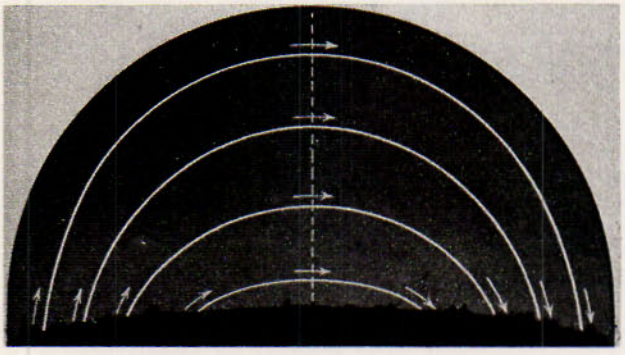


Fig. 61 — APPARENT MOVEMENT OF THE SKY: LOOKING SOUTH

seen by an observer in the northern hemisphere of the Earth, describe circles which are increasingly cut off by the southern horizon as they approach the south pole. And below a limit diametrically opposite to the northern circumpolar zone, there is a corresponding circumpolar zone where the stars are always invisible beneath the horizon. This zone is visible, however, all the time to people living at a southern latitude equal to the northern one we chose for our example above; in fact, it is their 'circumpolar zone'. Moreover, they never see those stars visible to observers in the northern hemisphere all the time.

So the equator is the only part of the world from which we can see the whole sky at some time during the twenty-four hours; the nearer we approach either of the poles, the smaller the portion becomes, and at the pole itself we can see only one half of the sky, the other remaining permanently hidden from our view.

Thus the sky seems to move in different ways when seen from different parts of the world. But except at the poles (where the stars move along circles parallel with the horizon) effects of perspective in the celestial sphere alter the shapes of these curves, which seem to move not only above but also around us. This is more easily seen when we take a long-exposure photograph of various parts of the sky: the diurnal movement causes the star images to trail on the plate, giving a vivid impression of the rotation of the celestial sphere (Figs. 62–7).

Whatever its path may be, whether high or low in the sky, a star reaches its highest point half way between rising and setting, when it crosses the meridian. This is the name given to the line joining the poles and zenith of any given place, and it defines the north-south direction.

For an observer in the northern hemisphere with his back turned towards the north, the stars move so as to rise on his left (in the east) and set on his right. For an observer in the southern hemisphere with his back towards the south, the stars will travel in the opposite direction, rising on his right and setting on his left.

If we observe the circumpolar zone from somewhere between the equator and the pole, the stars will be seen to cross the meridian twice: once above the pole, and, 12 hours later, from the opposite direction beneath the pole (Fig. 68). We call these superior and inferior culmination respectively.

The Earth's orbit: the ecliptic. — Knowing the general consequences of the Earth's rotation, the main one being the alternation of night and day, we must now study the effects of the revolution or orbital movement of our planet round the Sun. Many observed variations, e. g. the changing lengths of night and day, result from a combination of these two movements acting simultaneously.

Like all planetary orbits, the Earth's is an ellipse, with an eccentricity of 0.01674, and if we drew it exactly to scale, we should scarcely be able to distinguish it from a circle: we must therefore exaggerate it in Fig. 69. The first of the effects it has is that the distance between the Earth and Sun varies at various times of the year. The significance of the figure given for the eccentricity above will now become apparent.

At perihelion the Earth is 147 million kilometres from the Sun (approximately 91½ million miles), but at aphelion it is 152 million kilometres away (approximately 94½ million miles). The mean of these two, which we generally use when comparing planetary distances in the Solar System and which we call the Astronomical Unit, is 149 million kilometres or approximately 93 million miles.

The magnitude of these distances can be illustrated in the following way: suppose it were possible for an aircraft to fly from the Earth to the Sun at a constant speed of 1,000 m. p. h. To reach its destination, it would have to travel non-stop for roughly *ten years*. This is such a long time that it is unnecessary to try to be more precise by stating days and hours, etc. Indeed, it would be impossible to do so, for the distance between the Sun and Earth is not known to that degree of accuracy. In the account of how astronomical distances are measured, we shall later show that methods are liable to certain slight errors because of their extreme delicacy and the large distances involved, and that in the figure given for the distance to the Sun there is likely to be an uncertainty of about 10,000 miles in round figures.

Owing to the variation in distance mentioned earlier, the Sun's disc must appear larger at some seasons than at others. These differences, however, are so small that they cannot be detected with the naked eye and it requires optical aid to tell us that the disc varies from 32′ 36″ at its largest to 31′ 28″ at its smallest. The first of these values naturally corresponds to the time when the Earth is at perihelion, about 2 or 3 January, and the second to aphelion 6 months later, about 1 or 2 July. So during the first half of the year the Sun's disc appears to diminish in size and increases during the second. We might be tempted to draw an apparently obvious but paradoxical conclusion that it is larger and therefore gives more heat in the northern winter than in the summer, the difference in radiation being about 7%, a considerable percentage. We shall resolve this difficulty in the section dealing with the seasons.

The orbit of the Earth is called the Ecliptic, because it is in this plane that eclipses occur.

Projected on to the celestial sphere, this plane appears as a large circle. An observer on the Sun would see the Earth make one journey round this circle during the course of a year; or from the point of view of an observer on the Earth, the Sun appears to move along this same

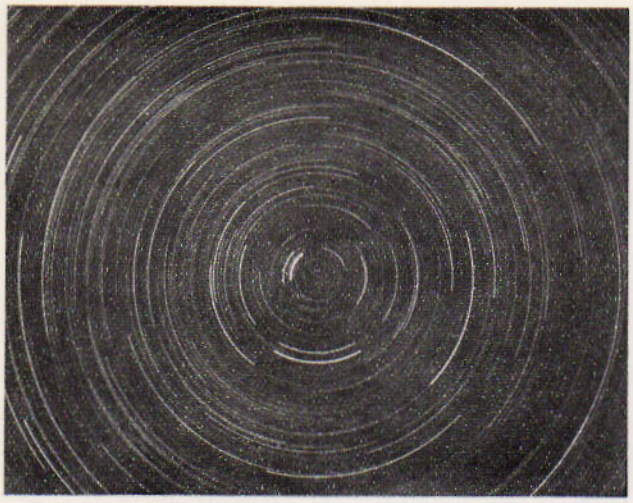


Fig. 62 — Apparent movement of the sky at the pole (Pole star)

Fig. 63 — Apparent movement of the sky at the northern horizon

Fig. 64 — Apparent movement of the sky at the north-eastern horizon

Fig. 65 — Apparent movement of the sky at the eastern horizon

Fig. 66 — Apparent movement of the sky at the south-eastern horizon

Fig. 67 — Apparent movement of the sky at the southern horizon (meridian)

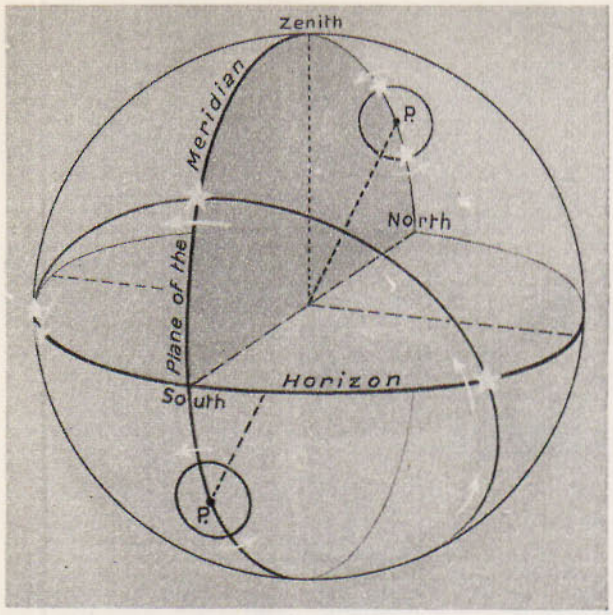


Fig. 68 — MERIDIAN TRANSIT (DUE SOUTH); SUPERIOR AND INFERIOR CULMINATIONS (DUE NORTH)

circle of the ecliptic, making one complete circuit, measured against the stars, during the year.

As we shall see, the ecliptic serves as a reference plane for defining the positions of other members of the Solar System.

The orbit defined above is purely theoretical. If the Earth were the only body to be considered, this would indeed be the path its centre describes around the Sun, but being accompanied by the Moon, our planet must maintain a condition of equilibrium with its satellite according to the laws of gravitation. The centre of gravity of the system is therefore no longer at the centre of the Earth, but about 1,000 km below the surface. Since it is the centre of gravity which describes the elliptical orbit and since the Earth moves about this point in order to maintain its condition of equilibrium with the Moon, and since the whole system then suffers further perturbations from the other planets, we may conclude that the Earth's actual path is extremely sinuous compared with is theoretical orbit.

This orbit is not invariable. The major axis of the ellipse makes one complete turn in 21,000 years; and the eccentricity itself undergoes a very slow variation, alternately tending to elongate the orbit or bring it closer to a circle. At present the eccentricity is diminishing and will attain its minimum value in about 24,000 years.

There are still further factors to be considered, but we can already see how complex the subject of the Earth's movement is. But for the most part, these changes are more of theoretical than practical interest, at least when we consider the minuteness of the effects they can produce during a single human lifetime, and they need not prevent us treating the principal movements in a somewhat simplified fashion, which is what we normally do.

The revolution of the Earth; aberration. — According to Kepler's second law, the law of areas, the Earth does not travel round the Sun at a uniform velocity; its velocity is constantly changing, being most rapid when the Earth is nearest the Sun. Its mean value is 29 k. p. s. or 107,000 k. p. h. (18 m. p. s. or 66,000 m. p.h.).

It will be noticed immediately that this speed far exceeds that of even our latest rockets, which can travel at about 25,000 m. p. h. Though our speed of revolution is so tremendous, we remain oblivious of the fact because we have no nearby standard of reference against which to measure it, such as we have in a train when we see the telegraph poles apparently speeding past the window.

Although we do not experience any feeling of movement, it is still possible to demonstrate it physically,

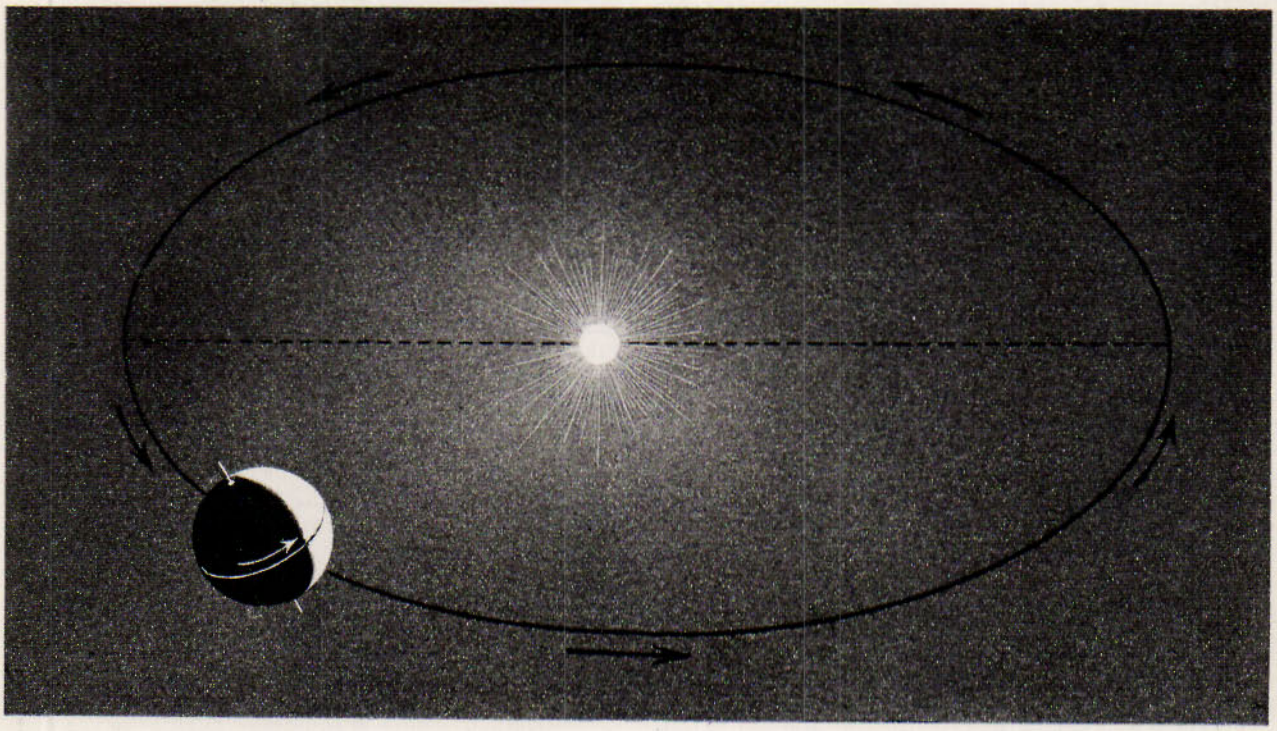


Fig. 69 — THE TWO MOVEMENTS OF THE EARTH: ROTATION AND REVOLUTION

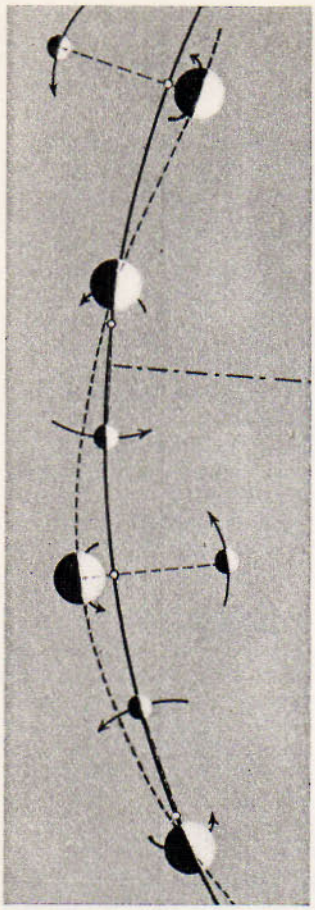


Fig. 72 — MOVEMENT OF EARTH AND THE MOON ABOUT THEIR COMMON CENTRE OF GRAVITY

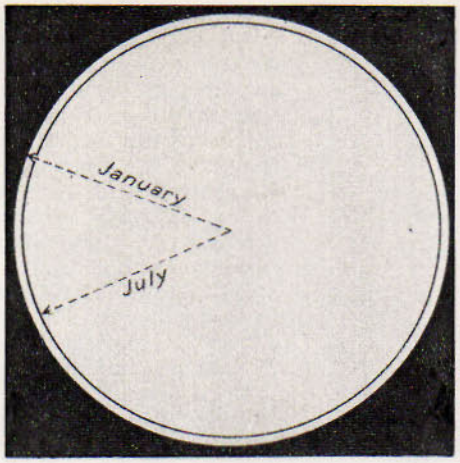


Fig. 70 — VARIATION IN THE SUN'S APPARENT DIAMETER DURING THE COURSE OF THE YEAR

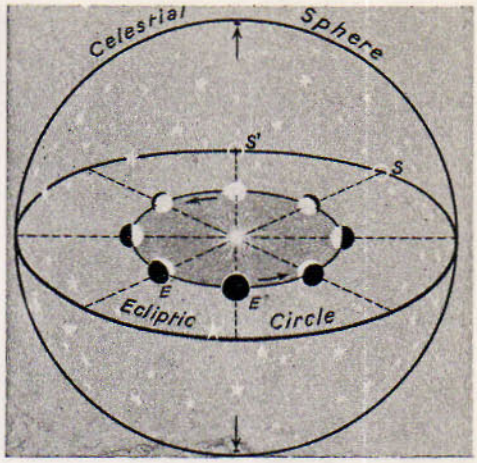


Fig. 71 — THE SUN TRACES OUT THE ECLIPTIC ON THE CELESTIAL SPHERE

just as Foucault's experiment with the pendulum proved that the Earth is rotating. This demonstration was made possible by the English astronomer Bradley (1693—1762) who, in 1727, the year of Newton's death, discovered the phenomenon known as the aberration of light.

To illustrate this phenomenon we shall use a familiar example taken from daily life. Let us imagine a shower of rain on a day which is calm enough for the raindrops to fall vertically. A man walking in this shower will think the drops are falling obliquely, since they will strike him in the face as he moves into them. If he has an umbrella, he will hold it forwards in the direction he is going, and the faster he walks, the more he will have to tilt his umbrella. We can observe the same thing from inside a train or car: the rain strikes the side windows and runs down at a slant according to the speed of the vehicle.

The explanation of this effect is very simple, and can be seen from Fig. 73. A drop of rain level with a man's eye at A will fall to B, which is directly underneath. But during the time it takes to fall, the man will have moved forward to A'; and seen from this new position, the point where the raindrop hits the ground will be behind him. With reference to the movement of his body, these successive positions during the fall of the raindrop have determined the apparent path A'B, the resultant of the two movements. The faster he goes, the more oblique the path of the rain, which is why this effect is most noticeable from a rapidly moving vehicle.

Now imagine the man standing on the deck of a ship which, for the purpose of our example, we will assume to be going round in circles on the sea (Fig. 74). As before, a shower of rain is falling vertically, but from the ship it will appear to be falling obliquely. When the ship turns north, the rain will seem to be coming from the north, and as it turns south, the rain will appear to come from the south. The same thing will occur at all points of the compass. This succession of different directions will therefore describe a cone about the vertical, and the greater the speed of the ship compared with the rain, the more open this cone will be.

Such is the principle of the aberration of light. We can replace the streaks of rain with parallel rays of light coming in from the distant stars: in this 'shower' of light, the Earth travels round the Sun in its almost circular orbit. In order to catch rays of light coming from the vertical, astronomers must incline their telescopes very slightly forwards in the direction of the Earth's motion, like the man with the umbrella. If we compare the circular path of the Earth with the circular course of the ship in the illustration, the direction of the inclination necessary to see the star will be changing constantly, and the result will be that instead of remaining absolutely fixed in space, the star will appear to describe a tiny circle like an image of the Earth's orbit. By very precise measurements Bradley was able to detect these movements and they give visible proof of the Earth's revolution round the Sun.

The apparent displacements of the stars caused by aberration are relatively small because the Earth travels much more slowly than light: in round figures, 30 k.p.s. against 300,000 k.p.s.; the corresponding angle in radians will be 30/300,000 or 1/10,000, or 20.5″ (1 radian = 206,264.8″). The most probable value of this angle, or constant of aberration, as derived from the very large number of observations made over the centuries, appears to be 20.48″. This information is very valuable, because, the velocity of light already calculated with considerable accuracy from laboratory measurements, we can use it to determine the mean velocity of the Earth in its orbit, and from this, the distance of our planet from the Sun.

Revolution of the Earth; the year. — The time the Earth takes to make one complete revolution round the Sun is called a year. We have already seen at what speed our world must travel in order to accomplish this immense journey of 590,000,000 miles.

However, in order to be able to define a 'complete revolution', we must have some standard of reference to use as a starting point. This reference point may be furnished by the line joining the Sun to a star situated at what may be considered an infinite distance in the plane of the movement. Alternatively, we can imagine the

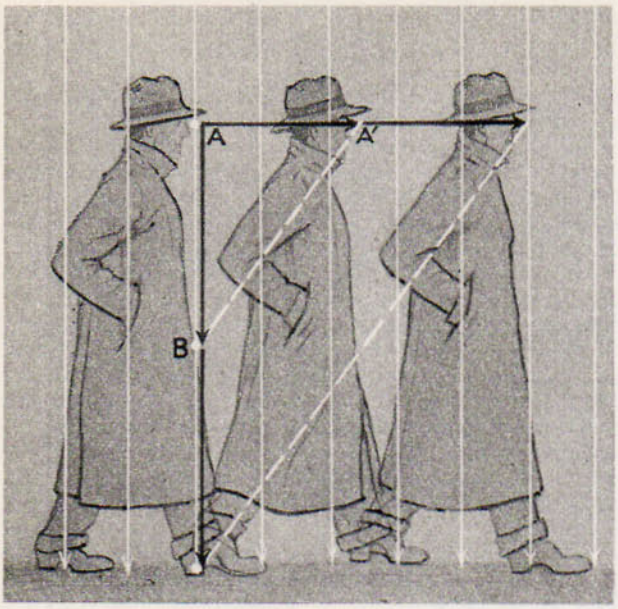


Fig. 73 — THE PRINCIPLE OF ABERRATION

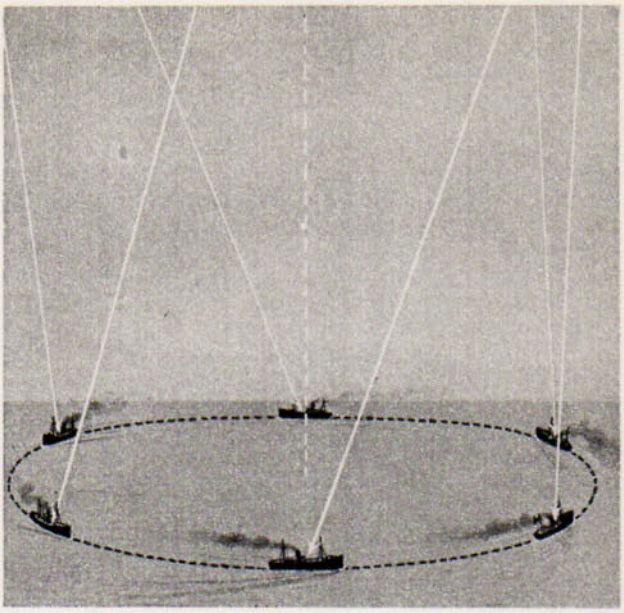


Fig. 74 — A FURTHER EXAMPLE OF THE PRINCIPLE OF ABERRATION

Earth in conjunction with a star as seen from the Sun. Every time the Earth crosses this line, it will have completed one revolution. The duration of this revolution period, or sidereal year, is 365d 6h 9m 9.5s, or, on the decimal system, 364.25631d.

But the revolution period can be measured not only with respect to the stars, but also in relation to the fixed points in its orbit, and, as an example, the tropical year is the time between two successive passages across the vernal equinox. We shall shortly learn what this term signifies and why the equinoctial point is not fixed but is constantly advancing in the direction opposed to the Earth's revolution. The Earth will therefore return to this point (which has, so to speak, come to meet it) after a time which is 20 minutes shorter than the sidereal year, i. e. 365d 5h 48m 45.975s or 365.24219879d. To all intents and purposes, the different seasons always start at the same period of the tropical year.

Combined effects of rotation and revolution. — The combination of these movements — rotation and revolution — will now make clear the reason for the difference which has already been mentioned between the Earth's sidereal rotation period and the civil day of 24 hours.

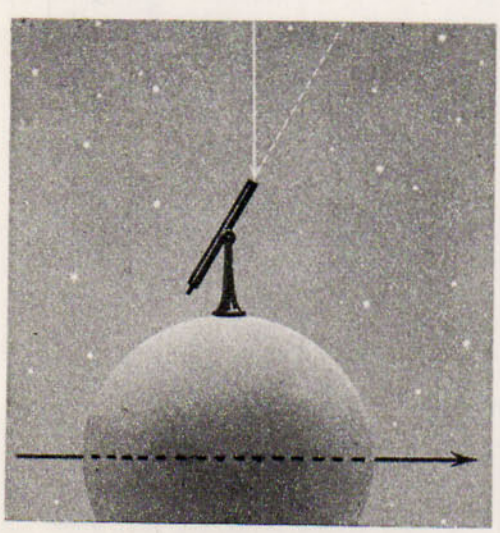


Fig. 75 — THE ABERRATION OF LIGHT

It might seem that the sidereal rotation period would be an ideal measure of the day, corresponding as it does to one turn of the globe on its axis. But since the Sun is the most important of the heavenly bodies from the point of view of humanity, man has since the very earliest times based his measurements on the Sun's apparent progress across the sky. This is how the difference arises.

We have seen that a star will return to the same point in the sky, the meridian for example, after one turn of the Earth about its axis. Suppose that we could observe both the star and the Sun in front of it, at the same time, or, alternatively, that the very distant star and the relatively close Sun could be seen in line. The Earth rotates and is also moving round the Sun. After exactly one turn, the point where we are stationed will have returned to face the star, because, being for all practical purposes at infinity, the star constitutes a fixed reference point. But in relation to the Sun the situation is very different: during the time the Earth takes to make one turn on its axis, the whole planet will have moved along its orbit by quite an appreciable distance, the line joining the Earth to the Sun acting rather like the spoke of a wheel. At the start of this turn the spoke pointed in the direction of the star and passed through the Sun. Now (Fig. 77) these two directions make an angle which exactly equals the arc through which the Earth has moved along its orbit. Thus, after passing the line to which the point A has returned after the course of one sidereal day, the Earth must rotate a little more for this point to be once again directly facing the Sun, or for the latter to have appeared to have made one complete turn round the sky. This extra rotation takes exactly 3m 56s, which is the figure mentioned earlier. The rotation period determined in this way, that is, the time elapsed between two consecutive transits of the Sun across the meridian, is called the true solar day.

However, this day cannot be used as a unit of time, since it is not of constant duration. In the above explanation we considered only the general orbital movement of

the Earth. As we have seen, this movement is not uniform: sometimes the Earth is accelerating, sometimes slowing down. Therefore, since the Earth does not cover the same lengths of arc in a given time at different seasons of the year, the extra amount of rotation which has to be added will be similarly variable, and this is reflected in the solar day. As these constantly changing conditions are unsuitable for the invariable measure of time we need, it has been found necessary to adopt the solution of the 'mean solar day'.

Let us take a first imaginary Sun, S′, around which the Earth moves uniformly, or, to conform more closely to appearances, which seems to move uniformly round the ecliptic. The plane of the Earth's equator being inclined to the ecliptic, the two intersect along a line which cuts the orbit at the two points we call the equinoxes (see below). We then imagine a second Sun, S″, travelling with uniform motion around the equator instead of the ecliptic; S″ starts from the vernal equinox at the same time as S′ and moves with the same speed that takes S′ round the ecliptic in a year. It is the movement of the second Sun which defines mean time. The unit of time therefore is the mean solar day, which is likewise used to define the tropical year.

The mean solar day is the time taken between two consecutive meridian passages of the mean sun for a given place. This moment is noon. Owing to the differences in the Earth's orbital velocity, true noon will coincide with mean noon only at certain times, and at others will be either ahead of or behind it. This means that the true Sun will sometimes cross the meridian before 'noon' as indicated by our clocks; sometimes it will be exactly punctual and at others will appear to be late. This variable interval, calculated for each day in the year, is called the equation of time.

Although time is measured by the meridian passage of the Sun, it would be impracticable to take this moment as the beginning of the day. Therefore, the civil day begins at mean midnight and is divided into 24 hours. Midnight is either 24h or 0h (the date changing at that time).

We need not over-emphasise the fact that the time determined by the Sun is not the same for all parts of the globe at once. The rotation of our world, which is illuminated only on its sunward half, carries every point first into the illuminated portion and then into the dark half. Thus, while the Sun lies due south of one place and it is mid-day, the opposite side of the world will be in darkness, and it will be midnight according to the Sun; and vice versa.

Every place, then, ought to regulate its time for its own meridian. The use of this local time, which is the kind given by sundials, was once very widespread, and the hour varied from town to town. Although this way of reckoning has its advantages, since it corresponds to the duration of the day at each location, it is beset with difficulties in communications with the rest of the world. Think of two people living in neighbouring towns; they arrange to meet each other exactly half-way between the two towns on the stroke of mid-day. But they would first have to decide the meaning of 'midday': is it the noon of the rendezvous or the noon shown by either of their watches? Whatever they decided they could not reach an agreement without a somewhat laborious calculation which would have to take into account the difference in local time between each of the two towns, and the time it would take for the two people to reach the rendezvous.

This problem would be particularly acute in a large modern country with rapid communication, such as the United States, where utter chaos would reign if everybody used their own local time. Even with the adoption of standard time zones, it is still possible for a New Yorker to send a telegram to Los Angeles and find that it has arrived several hours 'before' it left. And with the development of fast air transport, the rapidity with which the transcontinental traveller has to adjust his watch is phenomenal.

The circumference of the Earth at the equator is divided into 360°, representing the 24 mean hours of one complete rotation. By a simple division, we see that each hour corresponds to an arc of 15°, and a minute to 15′ on the circumference. Thus mid-day will be an hour later for places situated on wide circles or circles of longitude passing through the poles at intervals of 15°; that is to say, if we travel round the Earth on a course parallel to the equator, every 15° of longitude travelled will mean an adjustment of one hour to our watches, the time being added or subtracted according to our direction of travel.

Inclination of the Earth's axis; precession, nutation. — The Earth's axis is inclined at an angle of 23° 27′

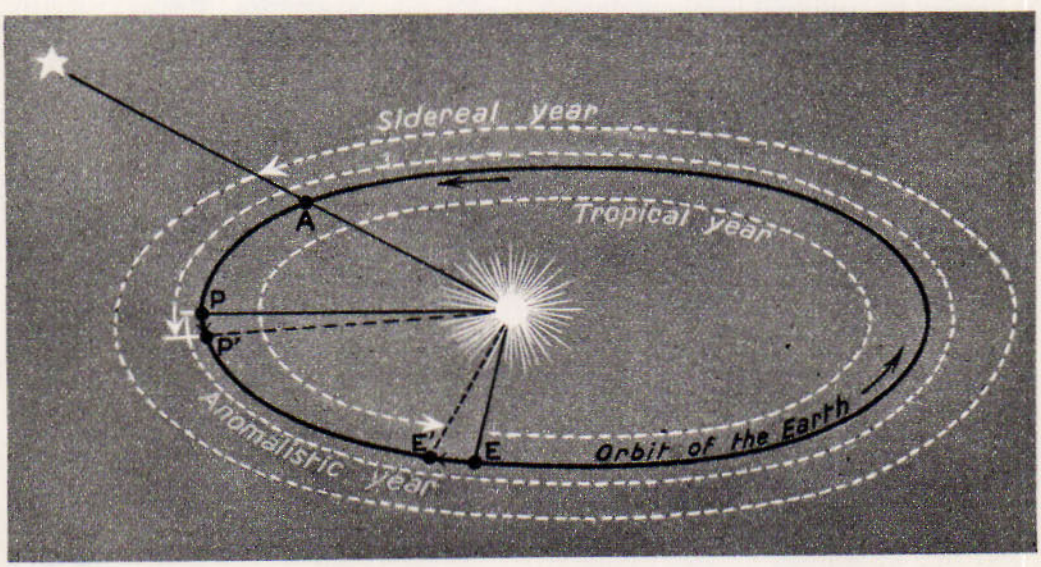


Fig. 76 — DIFFERENT SORTS OF YEAR

to the perpendicular to the plane of the orbit, or 66° 33′ to the orbital plane itself. As a result of this inclination, which always maintains a fixed direction in space, a given point on the Earth's surface receives the Sun's light at different angles through the year; this produces the phenomenon of the seasons and the unequal lengths of night and day for a given latitude.

To say that the Earth's axis always points in the same direction is an over-simplification we use to explain the mechanism of the seasons. In fact, the inclination is constantly undergoing very slight changes caused by a combination of various movements with various periods and magnitudes.

We have just said that the axis makes an angle of 66°33′ with the plane of the orbit; the largest of these other movements makes it describe a cone with the point at the centre of the Earth, and this cone traces on the celestial sphere a circle 23°27′ in radius around the pole of the ecliptic, or the point in the sky representing the vertical to the centre of the orbital plane.

This movement is similar to that of a spinning top which will balance in such a way that it describes a cone whose point coincides with that of the top. But although the cone traced by the top is similar to the cone traced by the Earth's axis, it is not absolutely valid to compare the celestial globe with a top. The difference is that the top balances on its point as it spins round, whereas in the case of the Earth this point is at the centre. What we have are in fact two cones meeting point to point.

This precessional movement is the result of the Earth's figure, which is not a perfect sphere, but flattened at the poles and bulging at the equator. This equatorial bulge introduces an element of asymmetry into the system; the attraction of the Sun, Moon and planets, which are all more or less in the same plane, tries to pull the equatorial bulge 'down' into the plane of the Solar System but this pull is resisted by the energy of the Earth's rotation. By way of compromise, therefore, the Earth's axis is caused to turn round in space.

This movement is extremely slow. In fact the time taken by the Earth's axis to trace out one cone, although slightly variable, is about 26,000 years. During the course of this long cycle, the celestial poles move along the circle thus traced out, and every star in their path becomes the pole star sooner or later.

For a star to be called the pole star does not imply that it must be exactly at the pole. It is sufficient if the star is so close to the pole that it appears stationary to the eye in comparison with the other stars moving round it. This is the case with our present north-pole star, which lies at a distance of about 1° (about twice the diameter of the Moon) from the true celestial pole. The effect of precession is to bring it gradually nearer, but the two

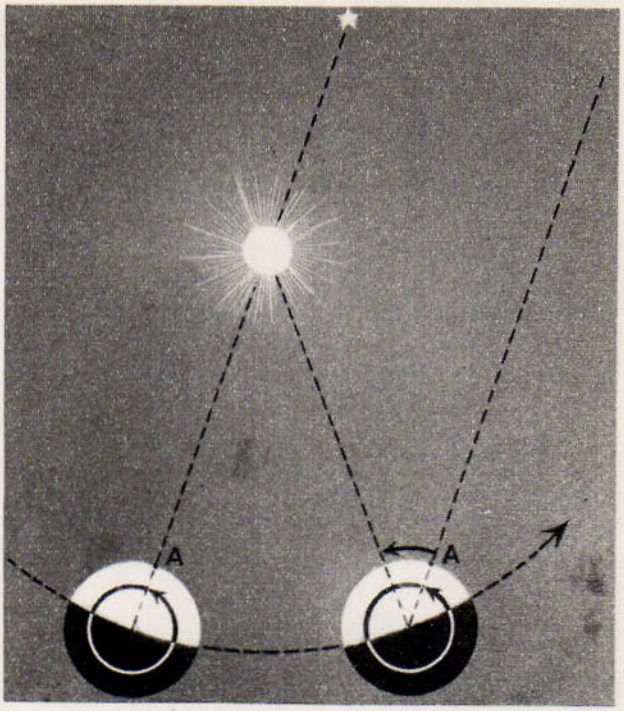


Fig. 77 — SIDEREAL AND SOLAR DAY

will never coincide exactly. At their nearest in A. D. 2100 they will still be 27.5′ apart.

In about 12,000 years time, Vega (α Lyrae), the brightest star in the northern hemisphere, will become the pole star, just as it was a little over 12,000 years ago. However, Vega will then be about twice as far from the pole as our present pole star is today.

The complete cycle of our pole's wandering through the northern sky is given in Fig. 85, which shows those stars which have been at or near the pole since the dawn of human history: for example, at the time of the earliest astronomers of Egypt and Mesopotamia, it was Alpha Draconis. It also shows pole stars of the future. It will be noticed that for long periods there will be no bright star near the pole; this is the case at present in the southern hemisphere, and the south pole is unmarked by any object bright enough to catch the eye.

Let us now consider the angle between the plane of the Earth's equator and the plane of the ecliptic, which is called the obliquity of the ecliptic. The two opposite points which mark out the intersection of these planes on the celestial sphere are called the equinoxes, the Sun crossing each of them once a year. As a result of the

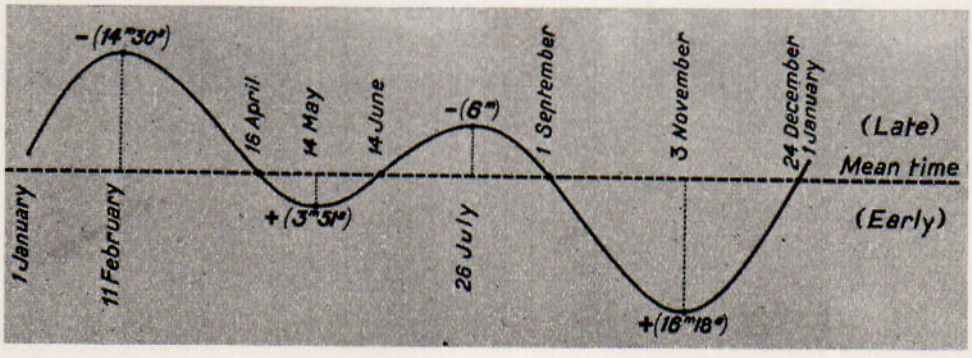


Fig. 78 — THE EQUATION OF TIME

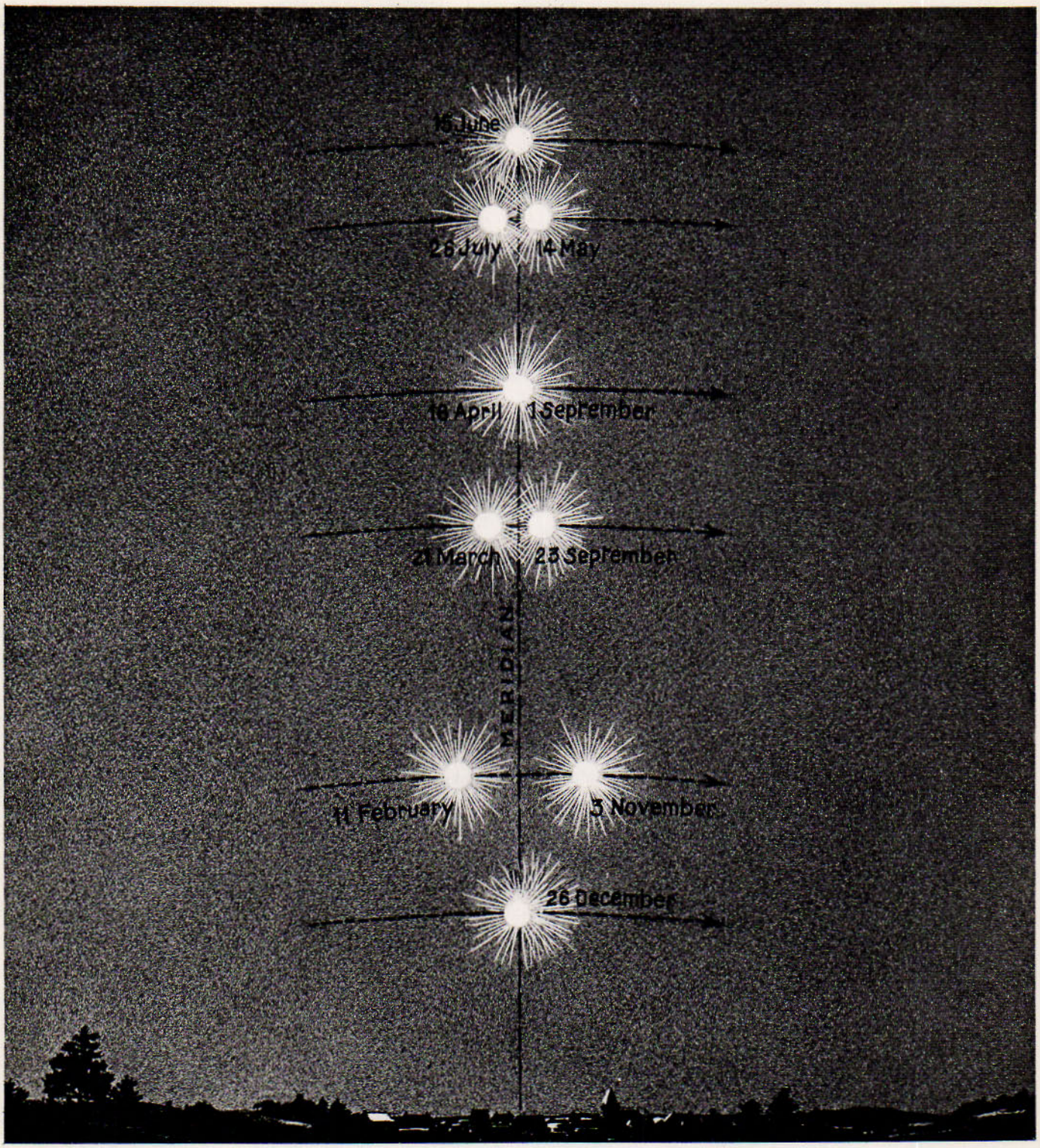


Fig. 79 — MEAN TIME AND SOLAR TIME. The Sun transits sometimes at noon, sometimes before or after.

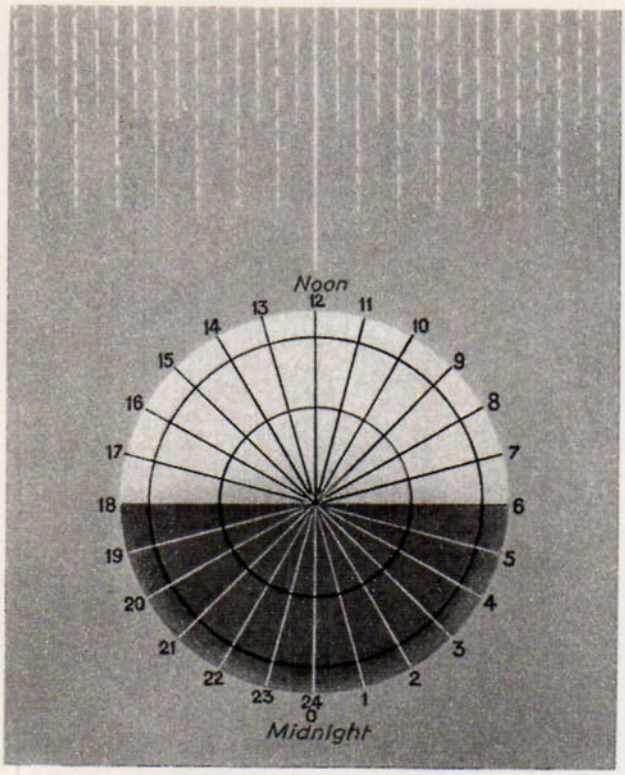


Fig. 80 — TIME AROUND THE EARTH AT HOURLY INTERVALS

conical movement of the Earth's axis, the plane of the equator, at right angles to the latter, shares the same precessional movement, and its orientation is constantly changing with respect to the ecliptic, but in the opposite direction to the way in which the Earth revolves. Thus at the same time as the poles describe their circles on the celestial sphere, the equinoxes seem to regress in relation to the stars, the annual value of the movement being 50.26″ of arc.

This explains why the precessional movement is also called the precession of the equinoxes.

As a consequence of this movement and in contrast to geographical positions, which are defined once and for all by latitude and longitude, the positions of the stars are gradually changing. This constant drift in the reference framework of co-ordinates must be allowed for in star catalogues, whose positions become appreciably incorrect after a few years. For this reason, they are always given for a stated epoch and the astronomer doing very accurate work has to calculate the correction for the particular time he is using them.

It follows from these facts that the stars visible from any one place will change over the years. Constellations visible at one period will gradually get higher above the horizon, while those on the opposite side of the sky will gradually disappear. The reverse process will then take place until the cycle begins again.

Having surveyed the general movement of the Earth's axis, we must now begin to realise that this movement does not take place in a perfectly regular manner.

In the first place, the obliquity of the ecliptic does not remain constant, but undergoes a very slow oscillation between 24° 36′ and 21° 59′. At present the angle is descreasing. The precessional cone does not in fact describe a perfect circle on the celestial sphere, but a curve which never quite closes, rather like a spiral slowly expanding and contracting in turn. Therefore, although the Earth's axis accomplishes a perfect circle with respect to

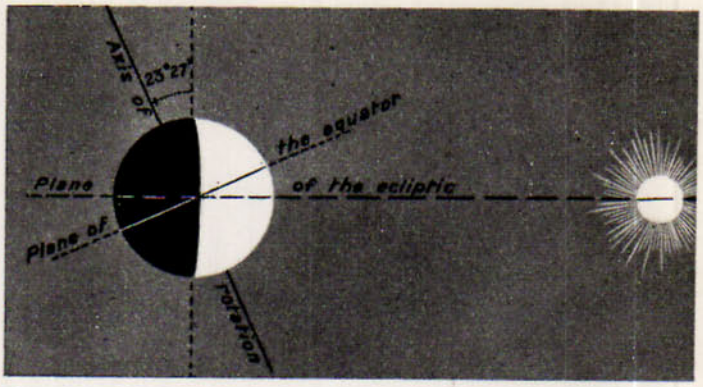


Fig. 81 — INCLINATION OF THE EARTH'S AXIS RELATIVE TO THE PLANE OF ITS ORBIT

the sky, it does not in fact come back to the same point in space.

In addition to this movement, there are others which cause further complications. In fact, when all the different movements are combined into one, the line the Earth's axis would trace out, supposing it could be equipped with the means to do so, would be extremely tortuous.

The most pronounced of these irregularities are due to the phenomenon of nutation. Nutation, too, is the result of a force exerted on the Earth's equatorial bulge, but this time attributable to the displacement of the Moon's orbit, the plane of which is constantly turning in space. It follows that this perturbing force, from a direction that is continuously changing, must cause the Earth's axis to deviate from the position it ought to occupy on the precessional cone, and it describes an ellipse 18″ long by 14″ wide in a period of approximately 18½ years, which is the revolution period of the plane of the Moon's orbit. The combination of this slight deviation of the axis and the precessional movement results in a wavy curve (sinusoid) as shown exaggerated for the sake of clarity in Fig. 87.

A second phenomenon of the same type, but much less pronounced than nutation, has a period of 6 months and is displayed as each hemisphere of the terrestrial globe is alternately turned towards and away from the Sun as the year progresses.

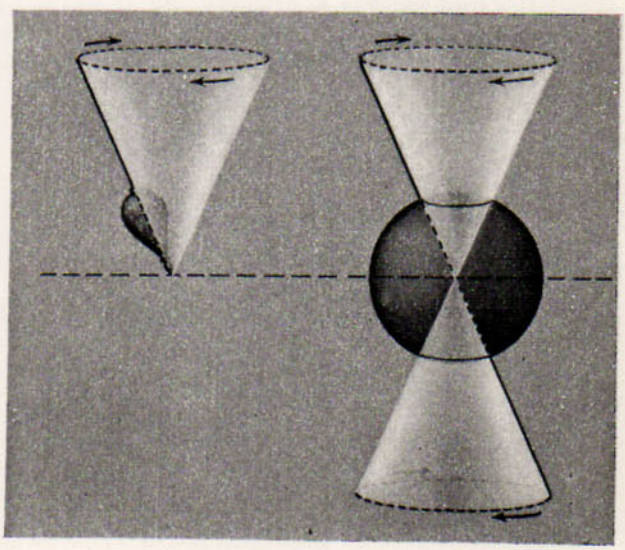


Fig. 82 — PRINCIPLE OF THE PRECESSION OF THE EQUINOXES

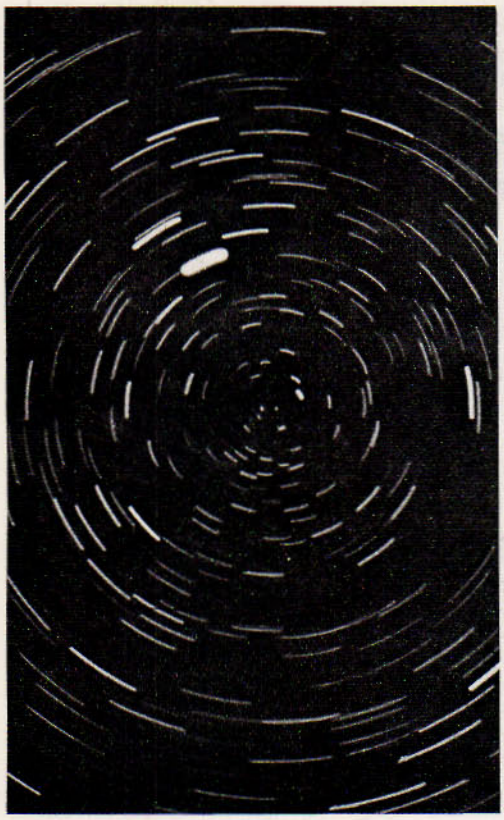


Fig. 83 — THE POLE STAR MAKES A CIRCLE AROUND THE CELESTIAL POLE

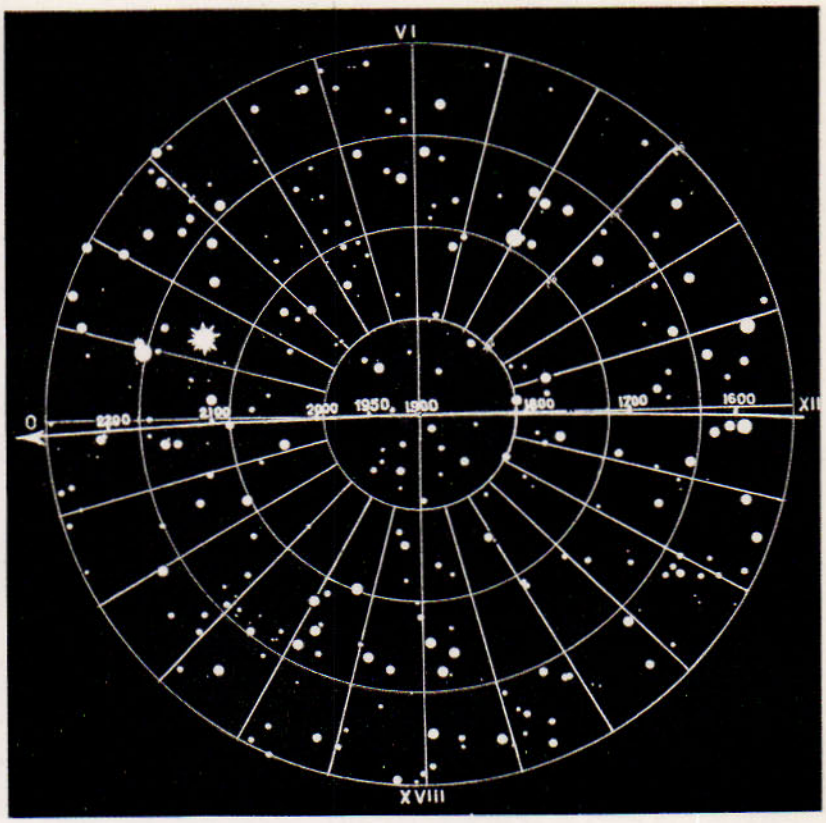


Fig. 84 — PATH OF THE CELESTIAL POLE FROM A.D. 1600 TO A.D. 2200

Yet a further perturbation is present, similar to nutation but quite insignificant, with a period of 14 days. It is caused by the obliquity of the Moon's orbit in relation to the Earth, which causes our satellite to exert its influence first on one hemisphere then on the other.

These then are the various perturbations which the imaginary axis of our world undergoes; and as if this were not enough, we may round off our account with an interesting piece of information discovered at the end of the last century. It is that the Earth's axis does not remain fixed in relation to the globe, like a skewer sticking through an orange, but is constantly shifting, even if only by a few yards. This means that the whole system of latitude and longitude on the Earth's surface is constantly and regularly changing, so that we cannot pinpoint our position with complete accuracy. We shall return to this interesting question in the chapter devoted to the Earth as a planet, because this effect can be attributed to the internal structure of our globe.

Equinoxes and solstices; the seasons. — There is no need to give a long explanation of the mechanism which produces the seasons as it can be understood from a simple diagram (Fig. 88). It will be seen that, since the Earth's axis maintains what is to all intents and purposes a fixed direction in space during the course of one revolution period, its orientation must change progressively with respect to the Sun. Since the plane of the equator is at an angle to the plane of the ecliptic, the Earth's rotation must take place obliquely; it follows that our planet's orientation to the Sun must be constantly changing. As may be seen from the diagram, there are four important positions in the Earth's orbit: the two solstices and the two equinoxes.

The Earth is at one of the solstices when the plane of its axial inclination passes through the Sun, at which time one pole of the globe is turned sunwards and the other away from it. A certain zone around one of these poles will remain in the illuminated portion whilst a corresponding zone at the other pole will have continuous darkness. These two positions correspond to the summer

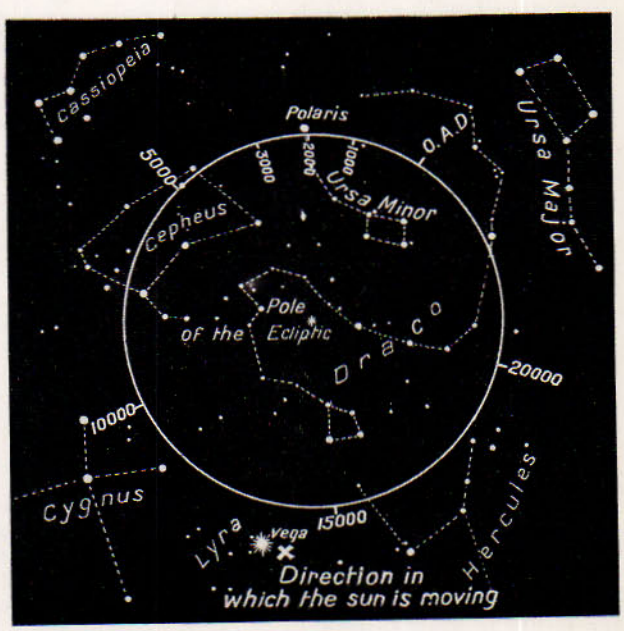


Fig. 85 — PATH OF THE POLE THROUGH ONE COMPLETE PRECESSIONAL CYCLE

and winter solstices respectively for each hemisphere, which is why the seasons change when we cross the equator.

The radius of the 'cap' which can remain in continuous darkness or daylight for 6 months of the year is necessarily equal to the angle between the axis of rotation and the perpendicular to the plane of the ecliptic. Since terrestrial latitudes are measured from the equator (3°) to the pole (90°), the parallels limiting these zones (called the Arctic and Antarctic circles), which must be at 23° 27′ from the pole, are at latitudes 66° 33′ north and south.

From these diagrams it can also be seen that the Sun

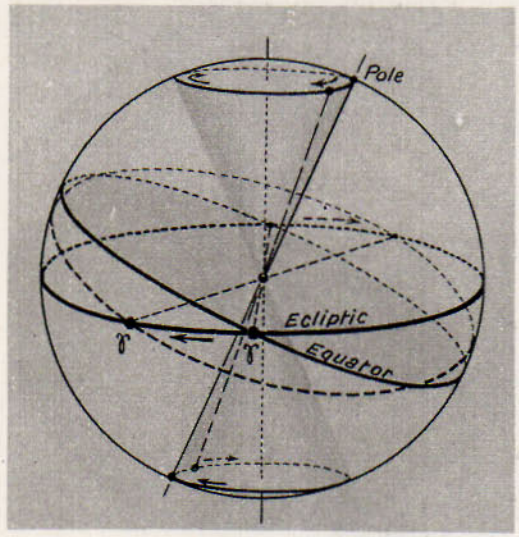


Fig. 86 — PRECESSION: RETROGRADE MOVEMENT OF THE FIRST POINT OF ARIES (♈)

is overhead at this time at any place on Earth with a latitude of 23° 27′ north of the equator in June and south in December. By virtue of the Earth's rotation all places on these parallels will pass successively throught this point and at that moment the Sun will be overhead. These latitude lines are called the Tropic of Cancer and the Tropic of Capricorn, because at that time the Sun is passing through those particular signs of the Zodiac.

Between the two solstices lie the equinoxes, where the Earth's axis no longer points towards the Sun but at right angles to it. The line of intersection between the plane of the equator and the plane of the ecliptic then passes through the Sun, which is overhead at noon on the equator, and both the poles then have the same conditions of illumination. One hemisphere, however, is at the vernal equinox whilst the other is at the autumnal equinox.

At some time during the year, the Sun is overhead at any point on or between the two tropics. Our globe can be divided into five zones (Fig. 89) according to the circumstances under which they receive the Sun's light: the two polar regions, inside the Arctic and Antarctic circles (lat. 66° 33′ north and south); the temperate zones lying between the latter and the tropics; and the torrid zone between the two tropics.

The year is therefore divided into four seasons determined by the two solstices and the two equinoxes: spring, lasting from the Sun's passage across the vernal equinox till it arrives at the summer solstice; summer, from the summer solstice to the autumnal equinox; autumn, from the autumnal equinox to the winter solstice; and winter, from the winter solstice to the vernal equinox again.

These four periods are not equal in length (Fig. 90) owing to the eccentricity of the Earth's orbit and to the angle of 10° which the line of the solstices makes with the major axis of the orbital ellipse. Thus by the law of equal areas, the Earth must take a different time to cover each of the above sections, from which the seasons in the northern hemisphere are as follows:

spring lasts 92.9 days;	summer lasts 93.7 days;
autumn lasts 89.6 days;	winter lasts 89 days.

These, moreover, are mean values only, because for various reasons that we need not discuss here the Earth does not reach the solstitial and equinoctial points at a set time. Thus it may be that a slight variation will mean that a season begins at 23h 59m instead of 0h 1m, thus changing the date. Such information is usually given each year in calendars and ephemerides, and so here we shall simply give the approximate date when the seasons begin; for the northern hemisphere:

spring begins	20 or 21 March;
summer begins	21 or 22 June;
autumn begins	23 September;
winter begins	22 December.

We have already mentioned that owing to the elliptical shape of its orbit, the Earth is further from the Sun in July than in January, receiving 7% more light and heat in January than July. As these dates more or less correspond with the solstices, these conditions should affect the seasonal characteristics: on the whole, the northern winter, occurring when the Earth is at its nearest to the Sun, tends to be less rigorous than the southern, which occurs when the Earth is receiving minimum light and heat. The northern summer is not so warm as the southern, when the radiation received is at its maximum.

The length of the day. — In this section, we shall see how the various positions that the Earth takes up in space affect conditions on the surface of the globe. Here again figures will be at the heart of the discussion, and we shall confine ourselves to essentials.

First, on account of the Sun's very great distance, we

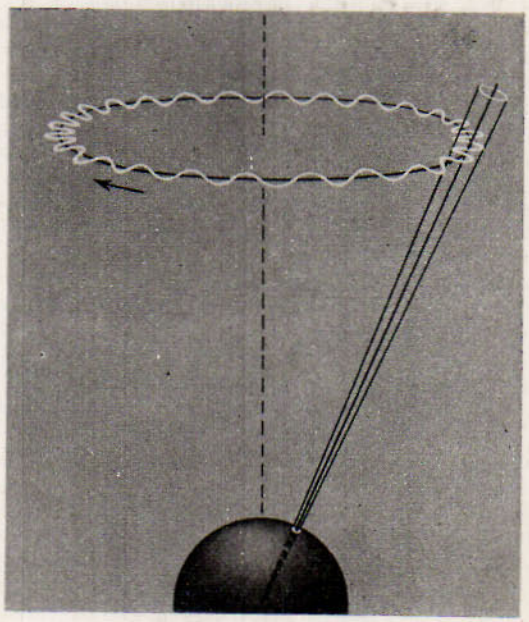


Fig. 87 — NUTATION

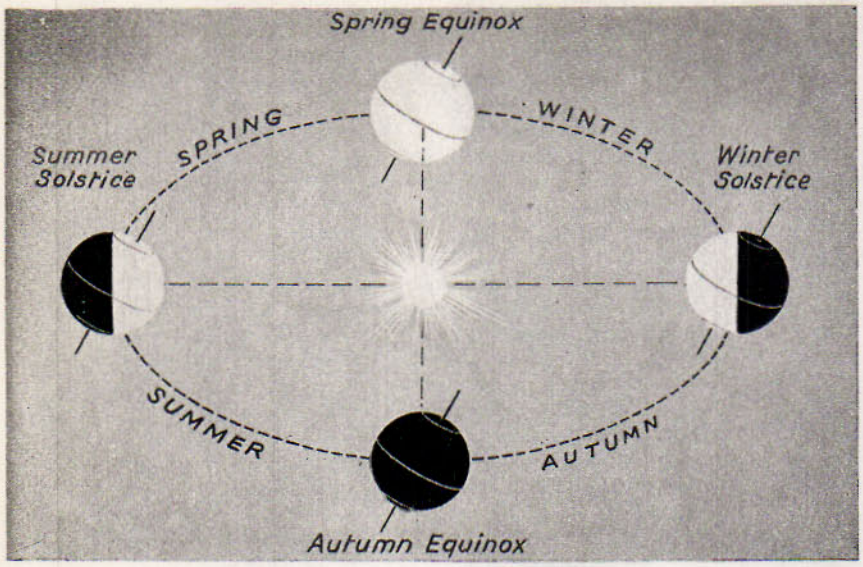


Fig. 88 — THE SEASONS (NORTHERN HEMISPHERE)

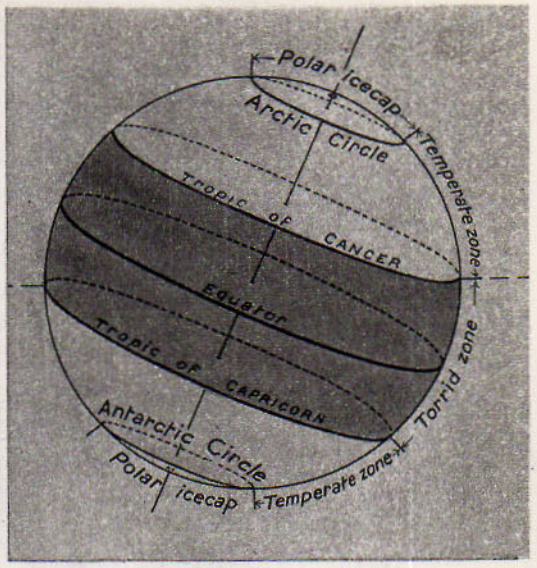


Fig. 89 — TROPICS AND POLAR CIRCLES. The five zones of the Earth

may consider the solar rays as being parallel when they reach the Earth. Consequently, the line separating day from night will cut the terrestrial globe along a great circle perpendicular to the direction of the Sun: this is called the circle of illumination or terminator. We shall meet this word again when we come to discuss lunar and planetary observations.

If the Earth's axis were perpendicular to the plane of its orbit, its projection would coincide with the terminator, which would pass through the poles; every point on the Earth's surface would spend half of one day in the illuminated part and half in the dark part of the globe. But since the axis is inclined at a considerable angle, and its orientation in relation to the terminator is constantly altering, conditions on the globe also change from one season to another.

Once again the reason lies in the orientation of the Earth's axis, which constantly makes an oblique angle with a fixed direction, in this case the terminator. And, as Fig. 91 shows, the proportion of the Earth's circumference which is in the illuminated portion varies considerably at different latitudes, the inequality being at its maximum at the time of the solstices. It may be noted moreover, that only the equator itself is exempt from these conditions, which affect all the rest of the Earth's surface. Whatever the planet's orientation with respect to the Sun, the equatorial circle is always cut into two equal halves by the terminator, and the lengths of night and day are practically the same there all the year round. Let us now see what happens at a different latitude round about the time of the solstice: for example, the northern summer.

As we leave the equator and approach the pole, which is turned towards the Sun and rotates full in its light (Fig. 92), the inequality of the division of latitude circles by the equator increases. Any point which stays longer in the illuminated portion of its rotational path will have a longer day than night. When we reach the Arctic circle, at latitude 66° 33′, the day becomes permanent, because from then on the whole Arctic region remains in the illuminated portion for the complete 24 hours. At the winter solstice these conditions are exactly reversed: the southern hemisphere imitates the behaviour of the northern but its summer is our winter, and vice versa.

Now let us consider the situation when the Earth has moved on to the equinox. The line joining the poles cuts the terminator, and Fig. 93 shows that conditions are the same for all latitudes, each making one half-turn in the light and one half-turn in the dark. This of course means that day and night are of equal duration at this season of the year.

Fig. 95 shows that the transition from one set of these conditions to the other occurs alternately in opposite directions. Fig. 95 is drawn for the northern hemisphere, but it can be applied to the southern simply by reversing it. The various positions shown bring out how the pole remains in darkness for one half of the year and in the illuminated portion for the other. The zone around the pole where night or day lasts for more than 24 hours can be seen to expand and contract, and this is mirrored in other latitudes by a similar lengthening and shortening of the periods of darkness. The difference becomes less pronounced as we approach the equator.

Because of the inclination of the Earth's axis, the Sun will take a different path with respect to the horizon at different times of the year, the angle varying according to the situation of the observer and the Earth's position

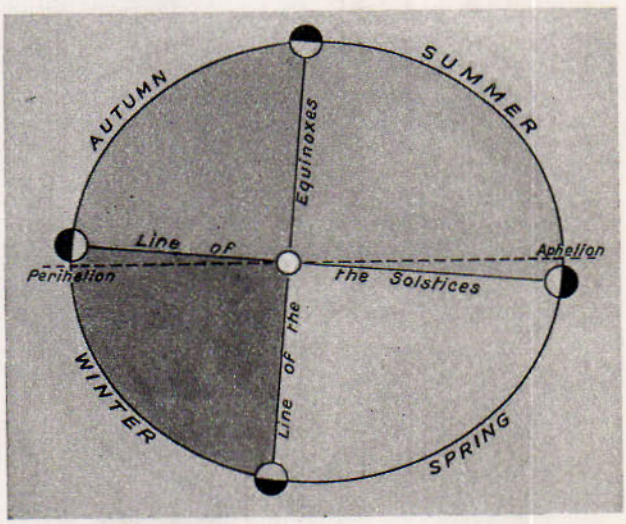


Fig. 90 — LINES OF THE SOLSTICES AND EQUINOXES. Unequal lengths of the seasons (northern hemisphere)

in its orbit. The constant change in the Earth's orientation with respect to the Sun means that the latter is continually changing its position in the sky, and it will move round with the celestial sphere along the parallel. Reference to Fig. 94 and to the examples already given should make quite clear what course the Sun takes in relation to the horizon at various times through the year.

The Sun's constantly changing celestial latitude, as we have just seen, is the result of its apparent journey round the ecliptic in the course of one year: as the Earth's equator is inclined at a considerable angle to the ecliptic, the Sun must cross it twice during the 12 months. This oscillation between one tropic and the other shows itself in the changing lengths of night and day, and we shall now explore this subject further.

For the sake of simplicity let us assume that the Sun moves along the ecliptic by an equal amount every day. This uniform displacement is exaggerated in Fig. 96, which shows very clearly that near the solstices the Sun's celestial latitude or declination (see later) changes very slowly, whilst near the equinoxes it changes very rapidly.

Compare these unequally variable conditions with those existing in the celestial sphere: it will be seen that for a fairly long period either side of the solstices the Sun will rise and set at very nearly the same points on the horizon for day after day and will follow very nearly the same path across the sky. The length of the day will seem to vary little. On the other hand, near the vernal and autumnal equinoxes the days lengthen out and the nights draw in very rapidly, for at these times the Sun's descent or ascent of the ecliptic is at its steepest. This change in latitude can be observed by noting the position where the Sun sets for a few evenings running.

The seasonal mechanisms already described, especially the change in the Sun's position in the sky when seen from another part of the globe, explain the reason for the spectacle of the 'midnight sun' (Fig. 101). This somewhat colourful expression simply means that in certain parts of the world the Sun is visible for the whole 24 hours of the day. As will be seen from the diagrams, such conditions occur only inside the polar circles, which form the boundary of the zone turning in continuous daylight from the summer solstice onwards. From these regions, the Sun is visible throughout the whole of its daily round,

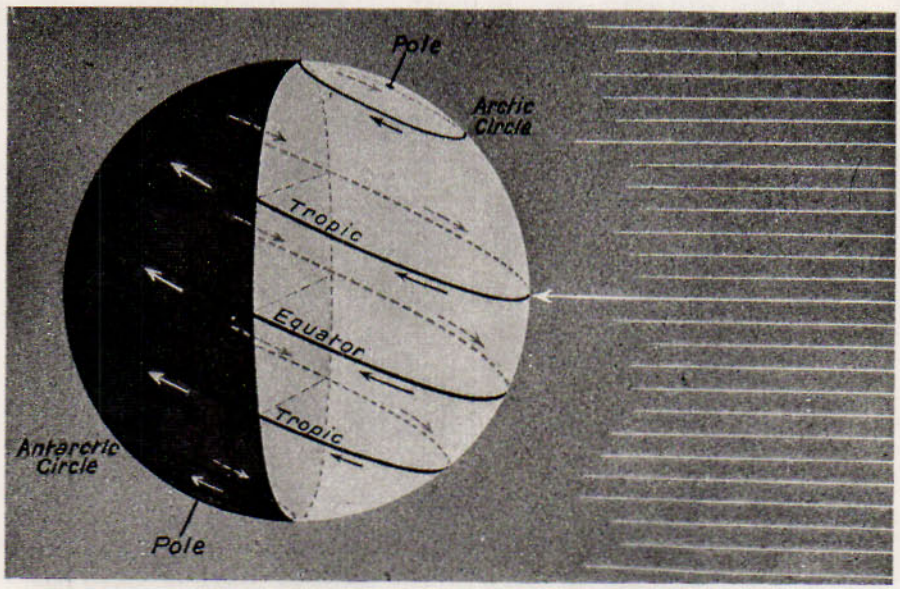


Fig. 91 — UNEQUAL LENGTHS OF DAY AND NIGHT

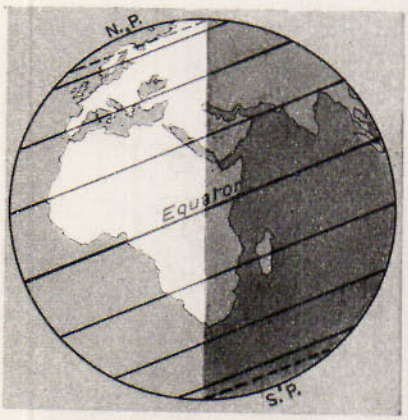


Fig. 92 — ILLUMINATION OF THE EARTH AT THE SOLSTICES. Night and day unequal in length.

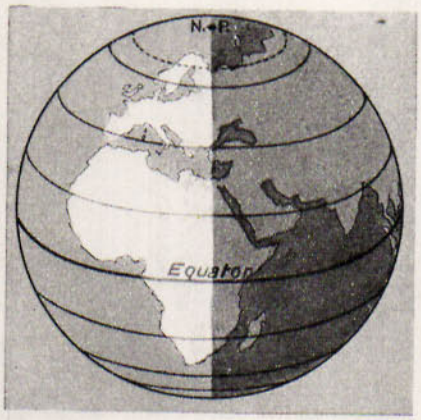


Fig. 93 — ILLUMINATION OF THE EARTH AT THE EQUINOXES. Night and day equal.

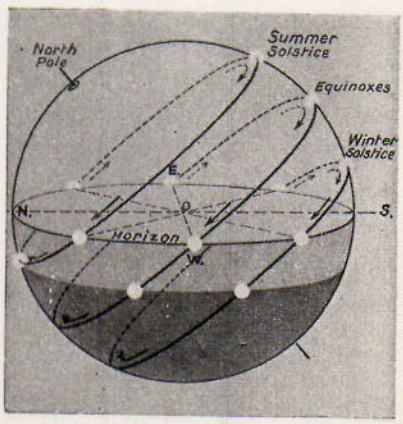


Fig. 94 — THE SUN'S APPARENT PATH ABOVE THE HORIZON

and, like a circumpolar star, first curves down towards the horizon, skims along it, then climbs up again, never having actually disappeared (Fig. 101). At the lowest point on its journey, that is when it is between the pole and the horizon, the Sun crosses the lower meridian: this marks the moment of midnight. For an observer exactly on the polar circle, the Sun will just touch the horizon at this moment. As we move into higher latitudes, the Sun's path becomes less oblique, and seen from the pole itself, the Sun travels parallel with the horizon.

On the other hand, when observed from a latitude just below the polar circle, the Sun actually disappears for a few moments beneath the horizon, and the duration of this disappearance increases as we approach the equator. Thus we return to the conditions already described, where day and night are of unequal length, the relationship changing with the seasons. However, there are a few points which should be explained a little more fully.

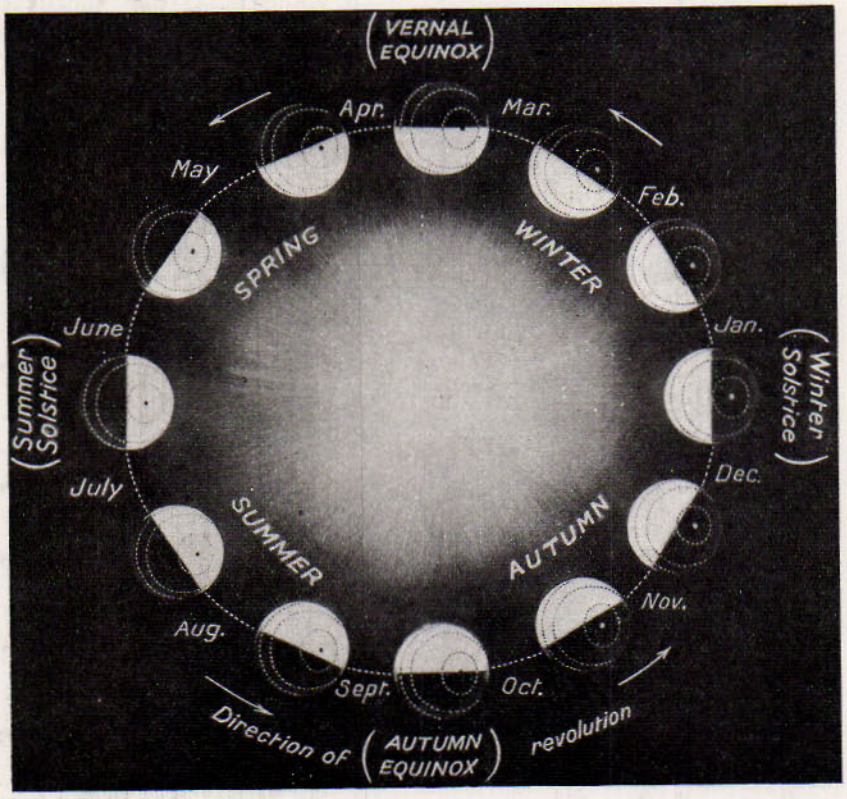


Fig. 95 — THE SEASONS

All these explanations of the unequal lengths of night and day are purely theoretical in so far as we have been attempting to deal only with the mechanism which produces them. In practice we must allow for the effects of our atmosphere, which considerably modifies the geometrical conclusions used in our explanations.

We have already seen what an important part is played

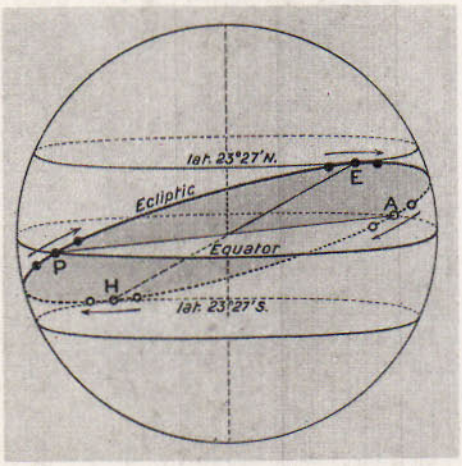


Fig. 96 — THE SUN'S MOVEMENT ON THE CELESTIAL SPHERE

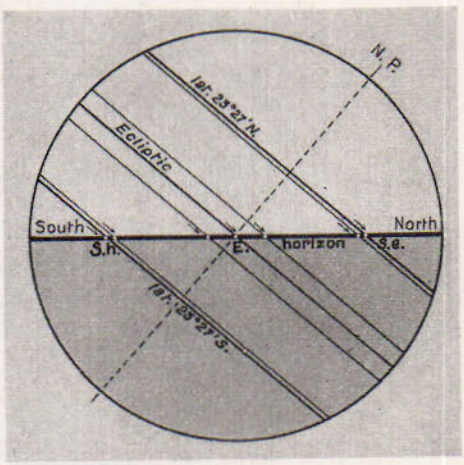


Fig. 97 — VARIATIONS IN THE SUNSET

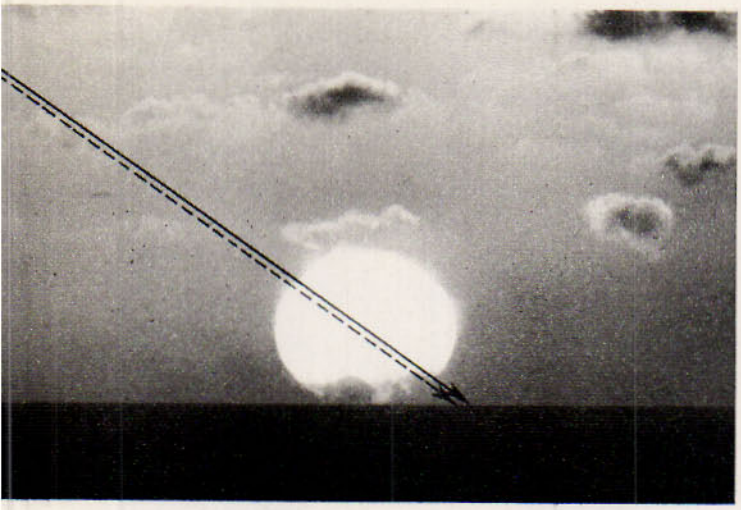


Fig. 98 — SUNSET IN SUMMER

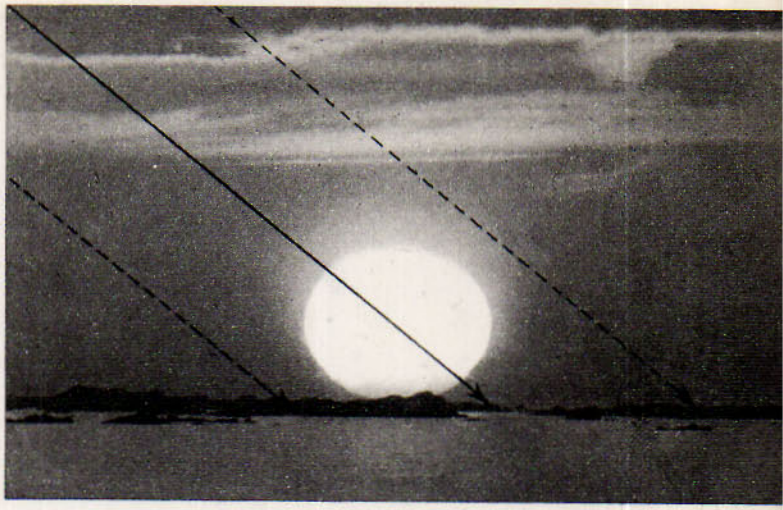


Fig. 99 — SUNSET IN WINTER

by our atmosphere in the diffusion of light, which obviates any sharp transition from day to night on the terminator. (We shall return to this point later.) Another effect we cannot ignore is the refraction of light. The most important point to notice about refraction is that it produces an appreciable difference between the direction in which a star rises and that in which it appears to. This means that a star can appear higher in the sky than it really is geometrically. At the horizon, refraction is at its maximum and its effect is so great that we can actually 'see' the Sun even though it may have sunk beneath the horizon (Fig. 103). Exactly the same phenomenon occurs in the morning, when we can be bathed in light before the Sun has really risen. Thus the actual duration of the day is several minutes longer than the theoretical duration measured as the interval between sunrise and sunset. As an example, at a place on the equator where day and night should be equal, the actual lengths are respectively 12h 5m and 11h 55m.

As has been shown, the midnight sun can be seen from the polar circles, which are defined as latitudes 66° 33′ north and south. But if we allow for refraction, which, at the horizon, can raise the elevation of a star by as much as 36′ above its true position, the Sun will be seen at midnight from a latitude 36′ further away from the pole. Consequently the latitude circle corresponding to 24-hour visibility of the Sun will not be 66° 33′, but 65° 59′. Conversely the night of the same duration, that is when the Sun never appears above the horizon (with which its path makes a tangent at mid-day) will not occur until latitude 67° 7′ us reached.

Thus it is inside these latitudes that real polar nights and days occur (Fig. 104). We must not, however, be tempted to over-generalise. In practice the polar circles

Fig. 100 — THE MIDNIGHT SUN AT TORNETRASK, SWEDEN (Courtesy Swedish Tourist Agency)

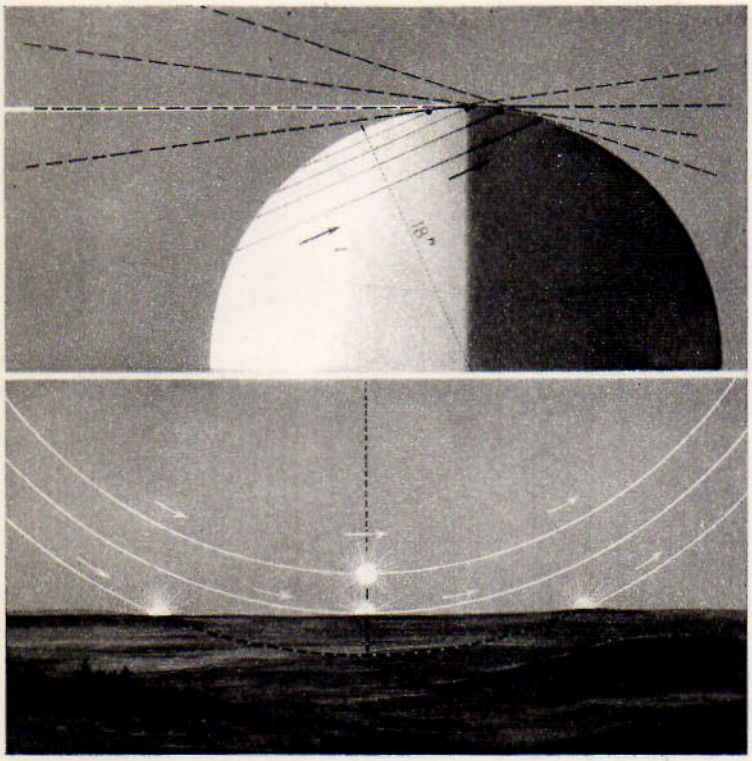


Fig. 101 — EXPLANATION OF THE PHENOMENON GENERALLY KNOWN AS THE 'MIDNIGHT SUN'

are often confused with the pole itself, and people tend to think that in the Arctic and Antarctic generally there is either six months' day or six months' night. This is only true in principle for the pole itself; elsewhere, the further a place is from the pole, the shorter these periods become.

Turning now to Fig. 102, we shall see in exaggerated fashion how the shift of the Earth's axis in relation to the terminator immediately following the solstices causes an increase or descrease in the area of the polar 'cap' remaining in the light or dark for 24 hours. Thus, strictly speaking, the Sun just fails to set on the actual day of the solstice when seen from the polar circle; on both the preceding and the following day, however, it does. For this latitude, there is theoretically only one day or night which can have a duration of 24 hours. In fact, because

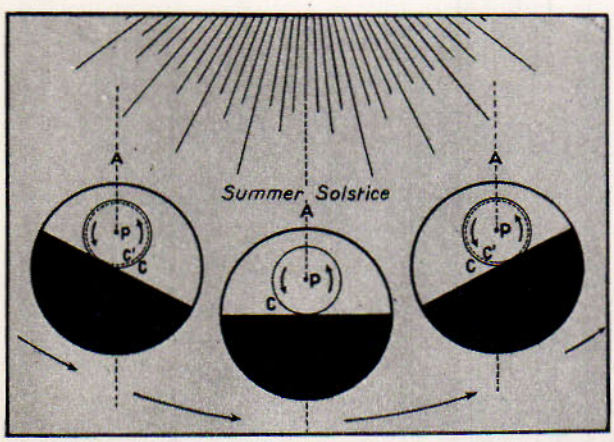


Fig. 102 — THE 24-HOUR DAY ON THE ARCTIC CIRCLE AT THE SUMMER SOLSTICE

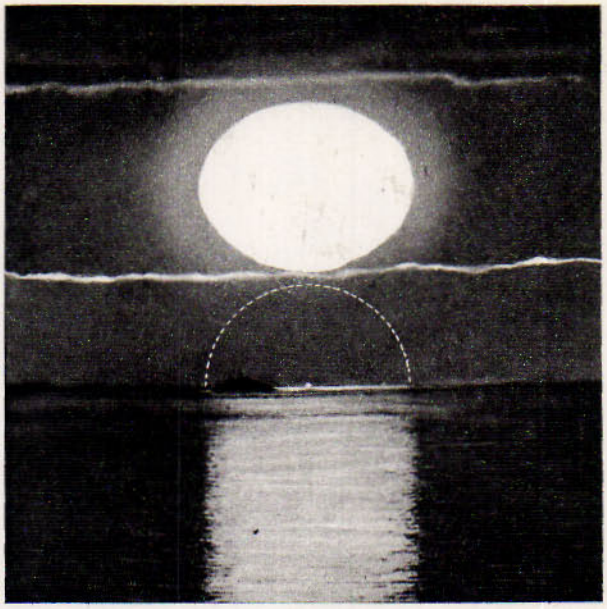


Fig. 103 — ATMOSPHERIC REFRACTION DELAYING THE SUNSET

of the slowness of the change in the Sun's celestial latitude at the time of the solstices, the Sun is visible at midnight for several days on end: but the topography of the horizon and the observer's altitude must be taken into account, for these also tend to produce slight changes in the Sun's visibility.

These principles, together with the accompanying diagrams, will help us to understand the progressive lengthening and shortening of the day as we go into higher and higher latitudes. It is at the poles themselves that the periods of night and day reach their maximum duration. These periods, moreover, are not identical for the two extremities of the Earth, for we must take into account the inequality of the seasons caused by the eccentricity of the Earth's orbit. These differences are summed up in the accompanying table.

LATITUDE NORTH	POLAR DAY	POLAR NIGHT	LATITUDE SOUTH	POLAR DAY	POLAR NIGHT
70°	70d	55d	70°	65d	59d
75°	107d	93d	75°	101d	99d
80°	137d	123d	80°	130d	130d
85°	163d	150d	85°	156d	158d
90°	189d	176d	90°	182d	183d

It may be as well to give some preliminary explanation of the terms used here. The expression 'polar day' is not ambiguous, because it means that the Sun makes its entire daily journey above the horizon. But 'polar night' does not mean exactly the opposite; for although, astronomically, we call 'night' the period between sunset and sunrise, that is to say the whole time the Sun spends beneath the horizon, this does not give a fair picture of what happens in nature. From the practical point of view, we enjoy an interruption or shortening of the period of complete darkness in the polar regions when, in theory, this ought to be continuous.

The reason for this is the phenomenon of the twilight, which was discussed in the opening pages of this chapter. It should be noted that there is both an evening and a morning twilight — the former fading after the Sun has

set, the other heralding the sunrise. Normally, the word 'twilight' refers to the evening twilight, and morning twilight is known as 'dawn-light'.

For most purposes it can be assumed that the twilight proper has ended when the Sun reaches an angle of 18° beneath the horizon, but from the above explanation of the apparent movement of the celestial sphere, and in particular of the Sun, it will be seen that, because of the varying slant of its path, the Sun can approach the horizon at different angles, therefore attaining its maximum distance below at different intervals after sunset.

In the equatorial regions of the Earth, dawn breaks very rapidly and night falls just as suddenly. This is because near the equator, the Earth's rotation carries the stars down towards the horizon at an angle which is very nearly vertical. This means that in the morning the Sun will rise up almost at a right angle from the horizon and sunrise will occur with maximum rapidity. In higher latitudes, however, the Sun approaches the horizon more obliquely and then sinks below it (Fig. 105). Thus in temperate regions the twilight, which is slow to appear and slow to fade, will move laterally along the horizon, since the arc which it seems to form is always centred on the Sun, and the latter, now invisible, is moving north or south with respect to the horizon.

Thus we can see the reason for our long summer evenings which seem to be immediately followed by the dawning of the new day. We have just seen that about the time of the summer solstices, within the Arctic circle, the Sun never comes nearer to setting than just skimming along the horizon at midnight, and so whenever it is less than 18° below the horizon its light will still be able to dispel the complete blackness of night. A simple calculation will show that the zone having not complete darkness but a persistence of twilight will reach 18° nearer the equator than the polar circle, that is to say 48° 33′ north or south. From this, it can be seen that London, say (latitude 51° 32′), and any other spot on the same parallel, has no true night in summer. Thus tourists visiting London from places further from the pole can enjoy a little of the polar night so dear to those who seek it in the far north of Europe.

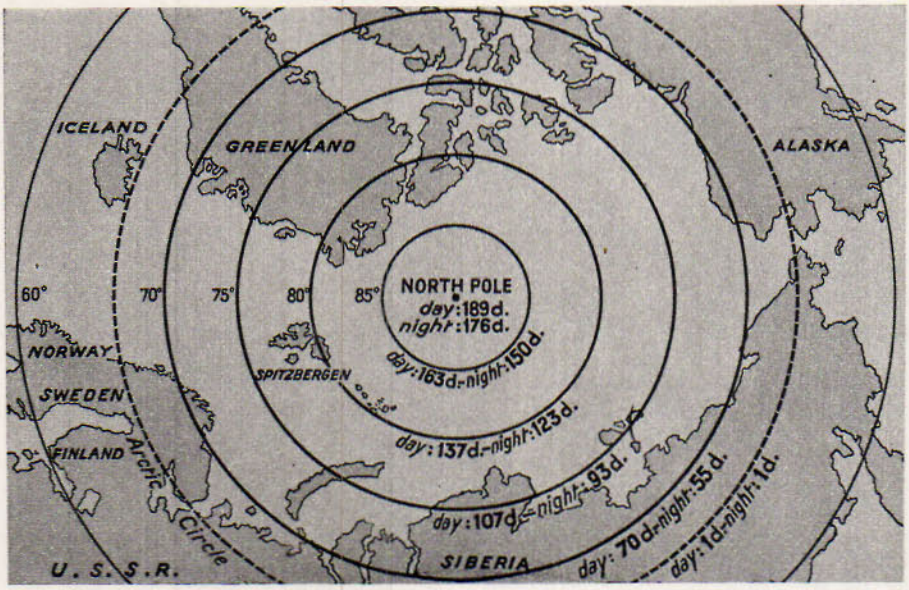


Fig. 104 — LENGTHS OF NIGHT AND DAY IN THE ARCTIC REGIONS

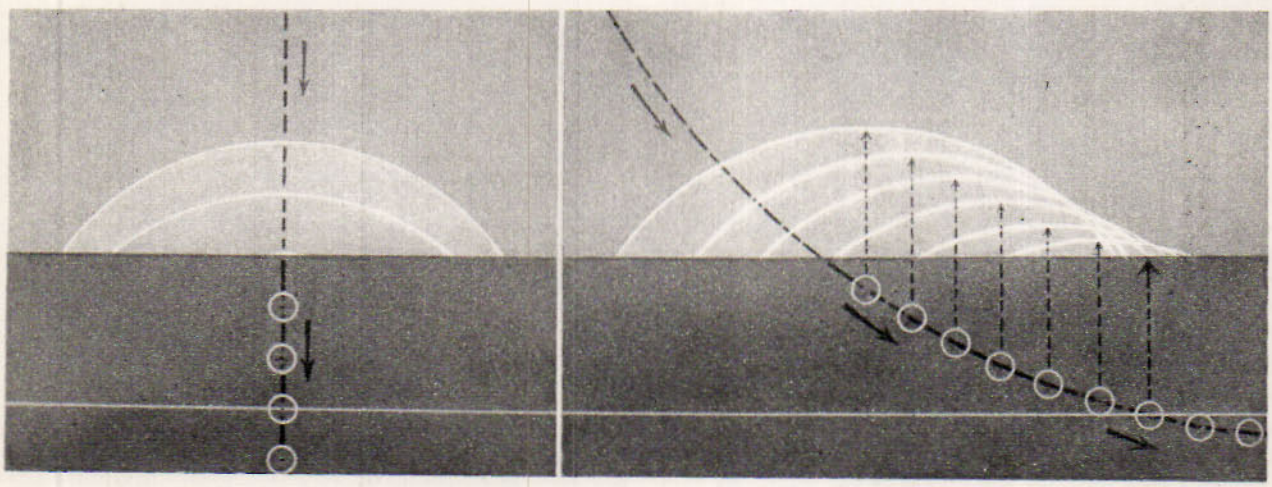


Fig. 105 — TWILIGHT AT THE EQUATOR (LEFT) AND IN HIGHER LATITUDES (RIGHT)

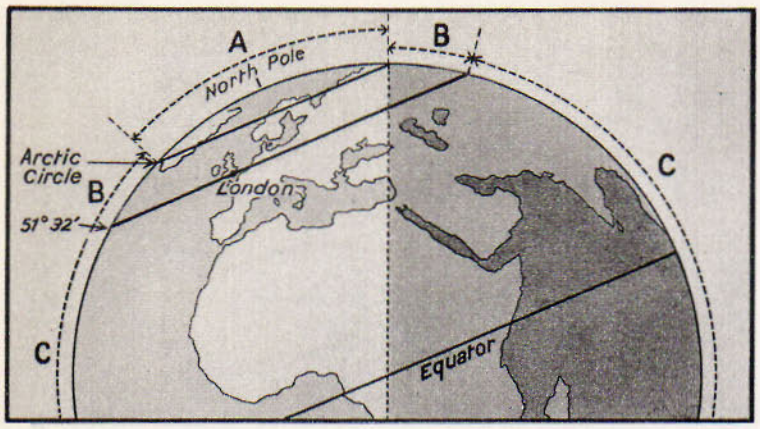


Fig. 106 — POLAR DAY AT THE SUMMER SOLSTICE, A, continuous daylight; B, night is not total; C, regular alternation of night and day.

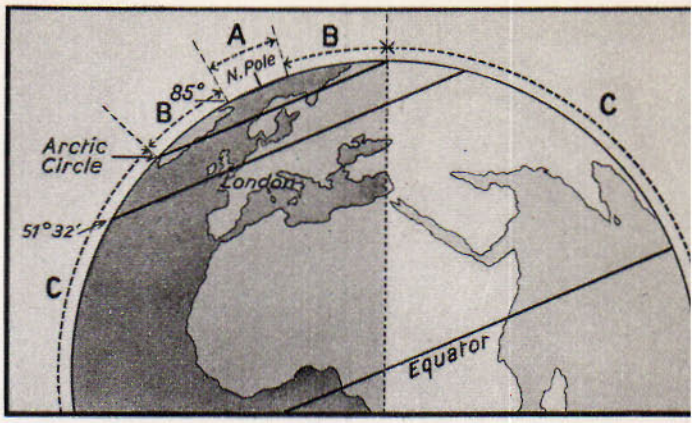


Fig. 107 — POLAR NIGHT AT THE WINTER SOLSTICE. A, night is total for 24 hours; B, night is interrupted by twilight; C, regular alternation of night and day.

As a matter of fact, this effect is a little more pronounced than it ought to be. The limit of twilight is determined by a calculation based simply on an assumed thickness of atmosphere, which, in fact, represents the thickest part. However, the atmosphere does not suddenly stop short but gradually becomes more and more rarefied up to some indefinite height. Twilight behaves in the same way. Although the direct illumination of the atmosphere by the Sun's rays ceases beyond a certain geometrical limit, the illuminated portion still reflects a diffused light for a certain distance beyond this limit, and so, extending a short distance above the twilight arc, by whose disappearance we judge the advent of night, there persists a diffuse glow which helps to shorten the period of complete darkness in our latitudes.

In addition to this astronomical twilight we also recognise a civil twilight, which lasts from sunset to when the Sun has reached an angle of 6° beneath the horizon; at about this time, if the night is clear, the brighter stars and planets begin to appear in the rapidly darkening sky and we can no longer see clearly around us. In contrast to this, astronomical twilight lasts until the faintest stars visible to the naked eye are visible in every quarter of the sky. As we have said, however, a faint glow persists even when the real twilight has faded.

Let us now consider what will happen in the polar regions when the Sun is constantly below the horizon.

First of all, we must reverse the above calculations. At the winter solstice, places of latitude 67° 7′, the limit beyond which the Sun remains below the horizon, have a 24-hour night. At mid-day in these latitudes, the Sun is just in contact with the horizon, and the nearer we approach the pole, the lower the Sun sinks beneath the horizon. In these conditions, the Sun is always less than 18° below the horizon; that is to say up to latitude 85°, the polar night is interrupted once every 24 hours by the appearance of twilight, but the further the observer is inside the polar circle, the shorter and less intense does this interruption become. At midnight on the polar circle itself, the night is therefore replaced by almost daylight conditions, since the Sun, being just below the horizon, makes the sky as bright as at sunset. Since this twilight appears well before midnight and fades away several hours later, the expression 'polar night' is an empty one for this latitude. As we increase our latitude, the duration and intensity of this interruption become progressively less, and from latitude 85° onwards the polar night is complete.

Thus, however simple it is in principle, the alternation of night and day is a more complex problem when we are trying to establish close divisions between nocturnal and daylight conditions for all points on the surface of our globe.

The Zodiac. — Because of the Earth's revolution the Sun appears to make one complete journey round the ecliptic in one year; the ecliptic cuts the equator at an angle of 23 ½° and the places where these two intersect constitute the equinoxes. For 8 ½° on either side of the ecliptic, the ancients recognized a zone which they called the Zodiac (Fig. 109); this is divided into twelve equal parts called signs, starting from the point occupied by the Sun at the vernal equinox. All the constellations crossed by the zone are called 'Zodiacal', and they form the background against which nearly all the planets move, since their orbits practically coincide with the ecliptic.

The Zodiacal constellations through which the Sun passes in turn were named as follows, starting from the vernal equinox:

Aries, the Ram	Libra, the Scales
Taurus, the Bull	Scorpio, the Scorpion
Gemini, the Twins	Sagittarius, the Archer
Cancer, the Crab	Capricornus, the Goat
Leo, the Lion	Aquarius, the Water Bearer
Virgo, the Virgin	Pisces, the Fish

This nomenclature, having lasted many centuries, is now out of date because the signs no longer correspond to the constellations after which they were named. This is because for centuries the precession of the equinoxes has been putting back the vernal equinox by 50.26″ each year: as we still make the Zodiac begin at the vernal equinox, and since this point has moved round in relation to the stars, the arbitrary divisions of the Zodiac have, so to say, 'slipped' in relation to the constellations and the two no longer coincide. Consequently, although the Sun still enters the *sign* of the Ram at the vernal equinox, it is in fact in the *constellation* of Pisces, which used to be last in the cycle.

And so, although almanacs still say that at the spring equinox the Sun is in the sign of the Ram, at the summer solstice in Cancer, at the autumn equinox in Libra and at the winter solstice in Capricornus, it must be remembered that these are the names of the signs, and they must not be confused with the constellations of the same name. The discrepancy is such that each sign now bears the

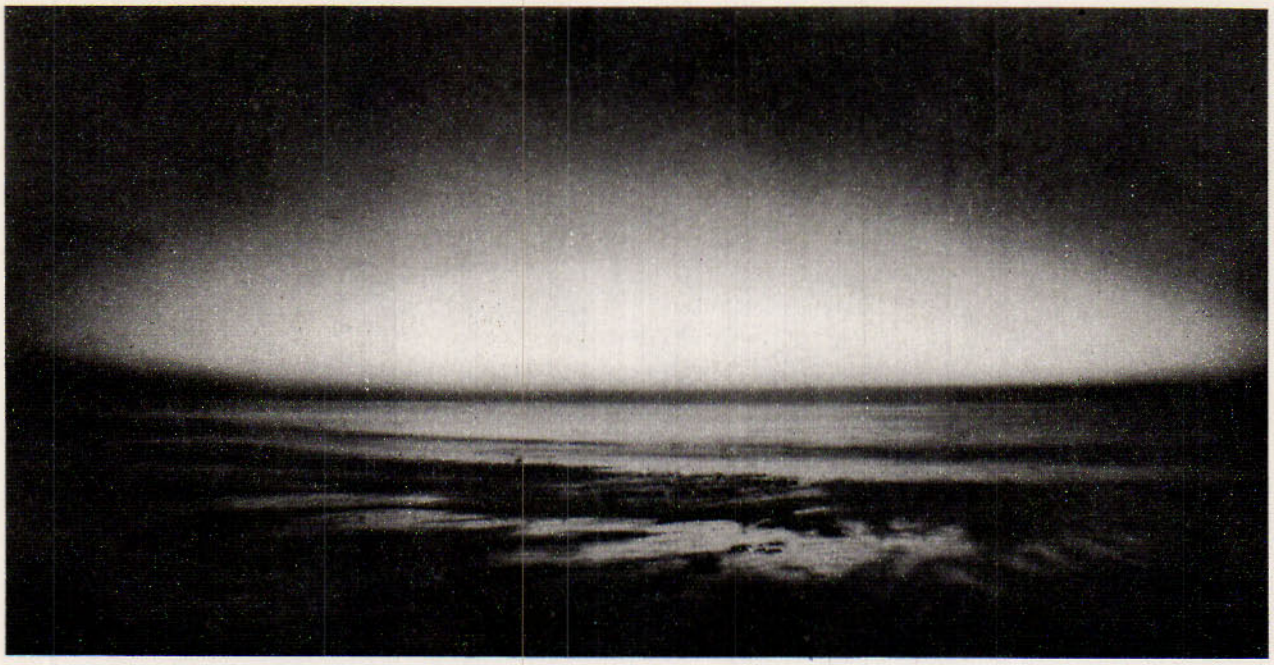


Fig. 108 — ASTRONOMICAL TWILIGHT

name of the constellation which preceded it in the original series.

Because of its movement along the ecliptic, the Sun will pass through every sign of the Zodiac during the course of the year. At the same time, almost the entire half of the sky that it occupies will be inaccessible to us because of the Sun's dazzling brilliance. This explains why we always see certain constellations in summer and others during the long winter nights.

Let us return to the mechanism which determines the difference between the solar and sidereal day. Timed by a clock, a star will return to the same position in the sky 3m 55s earlier from one day to the next, and this we can check by a very simple experiment. Examine the sky on a certain date, let us say 1 January, on the stroke of midnight, and notice which stars are just on the meridian, which are about to set and which are just rising. One rotation of the Earth will repeat these conditions 3m 55s earlier the next day, since we measure time by the solar day. But if we observe again exactly at midnight on 3 January, the sky will appear to have advanced a little compared with the night before: the stars which were just on the meridian will have moved off a little, those which were rising will be slightly higher in the sky, and those that were setting will have disappeared altogether.

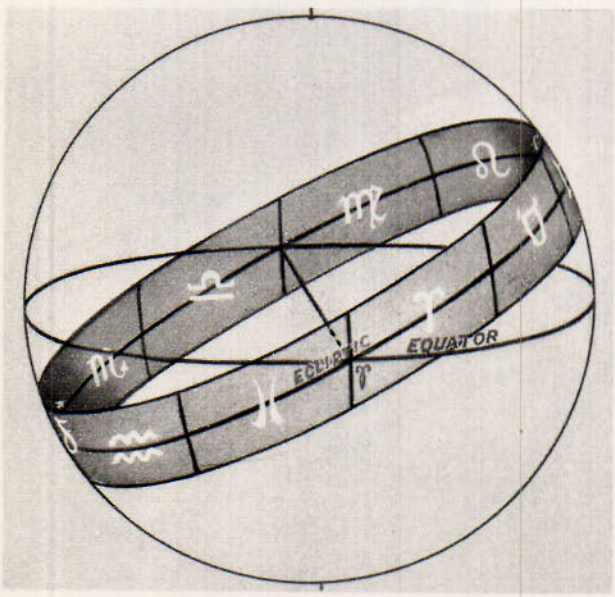


Fig. 109 — THE ZODIAC

Since, however, this advance is very slight and is therefore liable to be missed over one night, it is better to make one's observations at longer intervals; it will be found that over a month this daily advance will have added up to two hours, so that the sky will appear the same at midnight on 1 January, at 22h on 1 February, at 20h on 1 March and so on. On each of these dates we shall see a new portion of the sky on the meridian: this part of the sky is situated at 180° from the Sun.

We can now understand why the constellations visible in the middle of the night at one season seem to approach the Sun week by week until they finally disappear into the twilight glow. The exact opposite happens in the morning sky, where new constellations are constantly appearing out of the dawn.

The visibility of a given part of the sky, therefore, depends both on the time of day and the time of year. At intervals of 6 months at the same time of day, we find ourselves looking out in exactly opposite directions, the stars visible in winter being hidden by the Sun's light in summer, and vice versa.

In the application of these general principles we must take account of the observer's situation on the Earth's surface, and this is reflected in the orientation of the paths traced out by the diurnal movement of the stars. The changes in visibility we have discussed occur in those parts of the sky which we observe rising and setting. We may note that at the equator, the whole sky behaves in this way. From temperate latitudes there is a circumpolar zone which is constantly visible, although the shortness

of the night at about the time of the summer solstice reduces the possible observation time considerably.

As an example let us take the Great Bear, Ursa Major — the well-known constellation containing the Pointers which indicate the Pole Star. First we notice that the Great Bear is nearly always drawn in one position — the horizontal, which seems to make it easier to remember — and the line passing through the Pointers to the pole is then more or less vertical. Like the small hand of a clock, this arrangement makes one complete turn round the pole in 24 hours, but, as has already been shown, it will return to the vertical position a little earlier every evening.

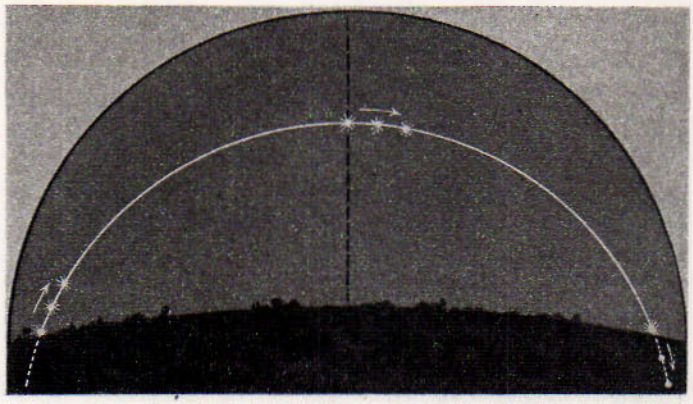


Fig. 110 — DAILY ADVANCE OF THE SKY

Alternatively, if we observe it at the same time tomorrow evening, it will have moved a little further from its present position (Fig. 114). A person unfamiliar with this behaviour may therefore be somewhat put out when trying to locate this constellation at some particular time, only to find that, at that moment, it is situated directly above the pole and therefore upside-down when compared with the way in which it is normally shown.

Therefore, when using the Great Bear to find the pole the observer must take this behaviour into account. To sum up, the Great Bear, like all circumpolar constellations, behaves rather like the hand of a clock, moving regularly but constantly gaining.

Celestial co-ordinates. — In order to check movements and determine positions on the celestial sphere, there must be some fundamental reference system of co-ordinates.

On the Earth's surface positions are found by means of latitude circles parallel with the equator, and great circles of longitude passing through the poles, originating from an arbitrary starting point.

The same procedure is applied to the imaginary celestial sphere: the Earth's equator, the other latitude circles and the poles are, as we have seen, projected on to the sky. The starting point of the celestial longitude circles is based not on an individual star but on the imaginary point where the planes of the equator and the ecliptic intersect, which is sometimes called the First Point of Aries and denoted by γ. This system of co-ordinates corresponds to the apparent movement of the sky due to the Earth's rotation, so that a star's diurnal path follows its parallel of latitude, and the great circle which fixes its position on this parallel crosses the meridian of a given place on the Earth's surface as a function of the sidereal rotation period.

These two sets of imaginary circles are not called by the same names as their geographical counterparts: in the sky latitude is called declination and longitude, Right Ascension, often shortened to 'dec.' and 'R. A.' respectively.

There is, however, another set of co-ordinates for which the terms longitude and latitude are retained. This is used when we measure not with reference to the equator and the axis of rotation, but from the plane of the ecliptic and its poles. This plane is determined by the Earth's revolution around the Sun. These co-ordinates are called 'ecliptic' whilst the others are 'equatorial'; the latter need not detain us, since they correspond to the way in which the celestial sphere seems to move and are those normally used in drawing up sky-charts. It may be added, though, that because of the precession of the equinoxes, these co-ordinates, which remain fixed with reference to the equator and the First Point of Aries, change in relation to the stars; conversely, it can be said that the positions of the stars slowly change against the framework of the co-ordinates.

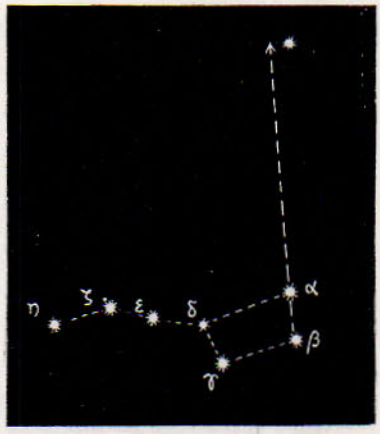


Fig. 111 — THE POLE STAR IN RELATION TO THE GREAT BEAR

The declination of a star, like the latitude of a town, is expressed in degrees, minutes and seconds. It is measured from the equator: north, the sign is positive, south is negative. In certain astronomical calculations, its declination from the equator is replaced by its distance from the pole, which is the complement of the former.

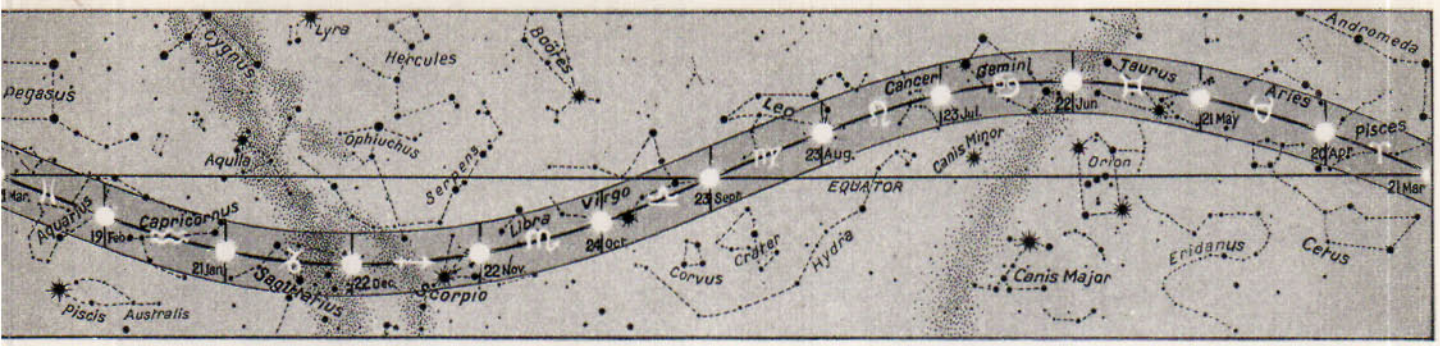


Fig. 112 — SIGNS OF THE ZODIAC COMPARED WITH THE CONSTELLATIONS

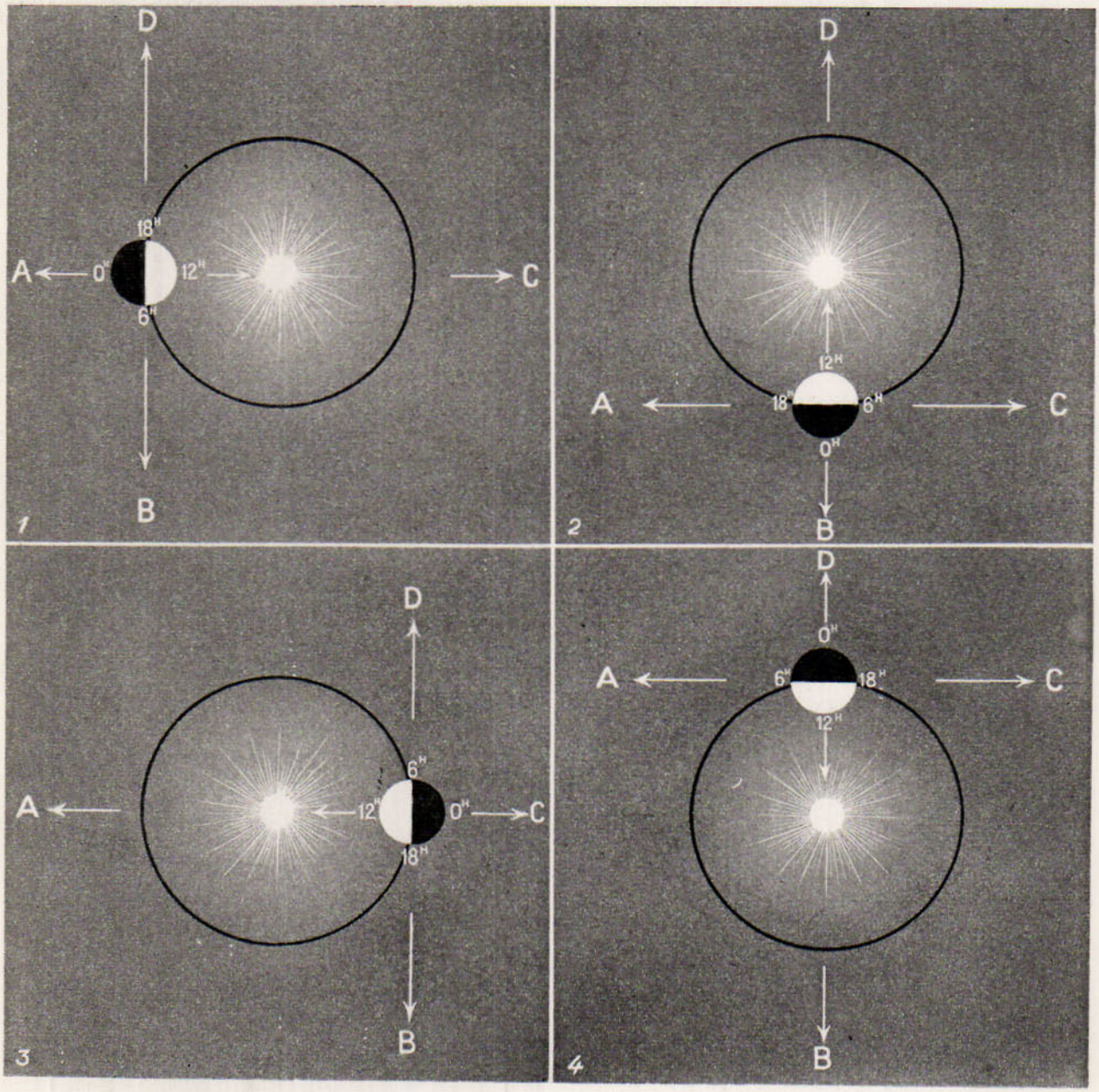


Fig. 113 — DIFFERENT PARTS OF THE CELESTIAL SPHERE ARE SEEN AT DIFFERENT TIMES OF YEAR

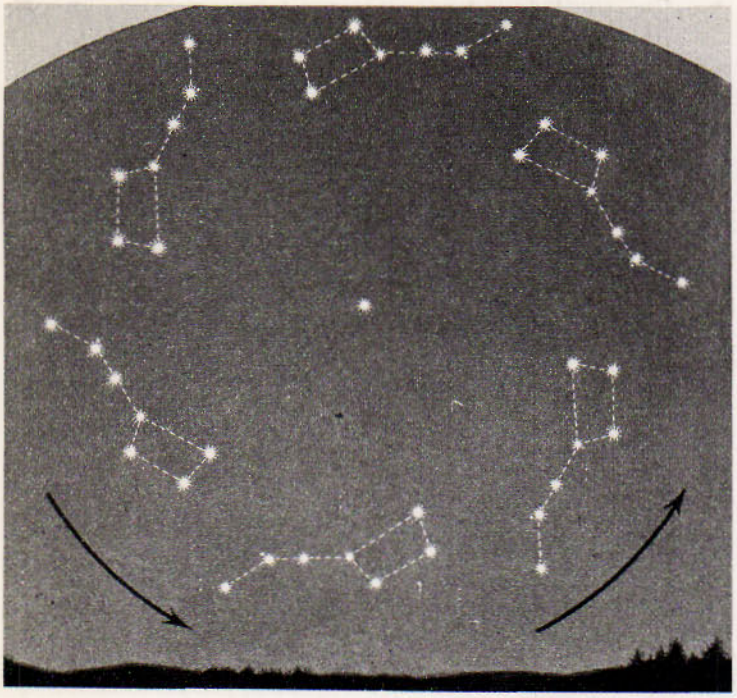


Fig. 114 — THE APPARENT MOVEMENT AROUND THE POLE OF THE GREAT BEAR

Right Ascension is differently expressed. The sky is divided longitudinally into 24 hourly divisions (Fig. 115), starting from the First Point of Aries, the vernal equinox, and going in the opposite direction to the diurnal rotation, which brings the hours of R. A. across the meridian of a given place in succession. The initial division is called oh, followed by 1h, 2h, 3h, and so on, the last being 23h. Any intermediate point is indicated in minutes and seconds. Thus a star may be at 2h 23m R. A. if the great circle passing through it follows the zero circle by this amount.

Each of the 24 sidereal hours into which the 360° of the sky is divided therefore has an angular value of 15°; each minute equals 15′ and each second 15″ of arc.

The length of an arc subtended by a given angle on the circumference of a circle will vary according to the radius of the circle on which it is marked. Therefore, the length of a 15° arc covered by a star on the celestial equator will be greater than the same angular distance covered near the pole. This can be seen from the long-exposure photographs showing the sky's rotation already given, with special reference to those of the circumpolar regions.

These differences, however, are not great enough to be readily appreciated with the naked eye, and we have to watch for some while before beginning to detect them. But there is one simple way of showing the movement of the sky to advantage, and this experience is perhaps the one which the young amateur starting in astronomy finds most striking. When a telescope is pointed towards a star and then clamped, the observer will see the star drifting quite rapidly through the instrument's field of view, and the greater the magnification used, the faster the star will drift. If, with the same optical arrangement, we compare the speeds of a star on the celestial equator and one near the pole, we shall be struck by the difference in the time it takes them to pass across the field of view: the equatorial star will fairly rush across, while the one near the pole will hardly seem to be moving at all.

Thus in precise astronomical work, visual or photographic, the astronomer must allow for these differences in movement. The determination of the co-ordinates which define the positions of heavenly bodies is one of the most fundamental branches of astronomy; therefore, in the final chapter we shall discuss the instruments and techniques used in this work.

The Earth's journey through space. — We have seen how far the Earth is governed by the various influences acting on it from the outside. Under the attraction of the Sun and planets, its movements lose the geometrical regularity which would result from a simple application of the laws of celestial mechanics. Sometimes accelerating, sometimes slowing down, rotating and wobbling on its axis, the Earth behaves rather like a soap-bubble floating through the air, on which the most imperceptible breath produces visible effects.

For the sake of simplicity, we have assumed up to now that the Sun is the fixed centre around which all the planets describe their orbits. For most purposes we can regard these orbits as closed curves, but when we come to consider the Sun as a member of the stellar system we call the Galaxy, the matter is quite different.

As we have seen, the Sun itself, like its countless companions, is moving through space in an orbit of its own together with its entire family of planets. This proper motion is the resultant of two very unequal component forces. One was recognised only a few decades ago, and is by far the more important; it is determined by the general rotation of the Galaxy (which we shall discuss in Chapter 16), implying a speed of about 270 k. p. s. for the Sun and nearby stars. The second, which has been known for almost two centuries, is much smaller, not exceeding 20 k. p. s., and is the result of the Sun's motion with respect to the neighbouring stars.

The second movement occurs obliquely to the plane of the Solar System. Since the Sun itself is moving while the Earth goes round it, the Earth actually moves along a spiral, or more correctly, helical path in space with reference to the surrounding stars, which for these purposes we can consider fixed. Seen from a great distance, the actual path of our world is rather like a gigantic corkscrew about 186 million miles in diameter and with a 'pitch' of about 590 million miles, the length of the Earth's annual journey with the Sun.

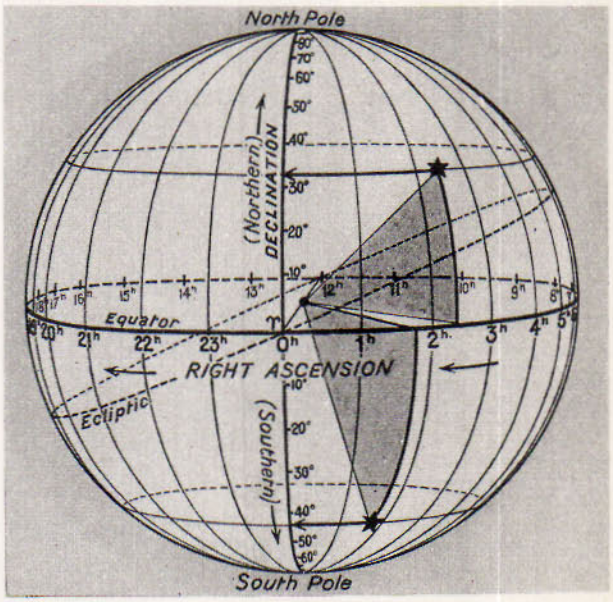


Fig. 115 — CELESTIAL CO-ORDINATES: RIGHT ASCENSION AND DECLINATION

Fig. 116 — APPARENT MOVEMENT OF THE SKY IN 5 MINUTES AT LATITUDE 55° (GREAT BEAR)

Fig. 117 — APPARENT MOVEMENT OF THE SKY IN 5 MINUTES AT THE EQUATOR (ORION)

We must, however, point out that in our normal lives we are quite unconscious of this colossal movement. The Moon and planets move together with the Earth just as if the whole system were stationary in space, and we can detect the movement of the system only by observing some comparatively fixed reference point outside it. For this purpose we may have recourse to the general background of stars to be found all around us. If the whole system is moving, we ought to be able to see the stars opening up in the direction in which we are going, and closing in behind us, like lamps as we walk along a street. These effects of perspective can be observed by making extremely delicate measurements of the proper motions of a large number of stars; this subject will be fully discussed later. It must be emphasised that these apparent movements are so minute that during the centuries men have been charting the skies, the changes produced in the positions of the stars have not been great enough to be noticeable by the naked eye, and for all practical purposes the stars might just as well be at rest in space.

Nevertheless, careful measurements made over the course of several centuries have enabled us to establish quite definitely that the Sun is moving in the general direction of the constellation Lyra, towards a point not far from the bright star Vega.

This movement is essentially relative. If, for example, we suppose that the Sun is moving in a straight line (and there are no grounds whatsoever for this assumption), it will never actually arrive at the constellation for which it is heading, for the very good reason that this does not exist! As we have already seen, a constellation is merely the appearance of a number of stars distributed in such a way that, although they are more or less in the same line of sight, the distances between them may be anything up to thousands of light-years. We cannot therefore talk about the 'distance of a constellation', only about the distances of its member stars. Even this does not mean that the Sun will ever reach one of them, for while the Sun is moving towards them, they are moving away in

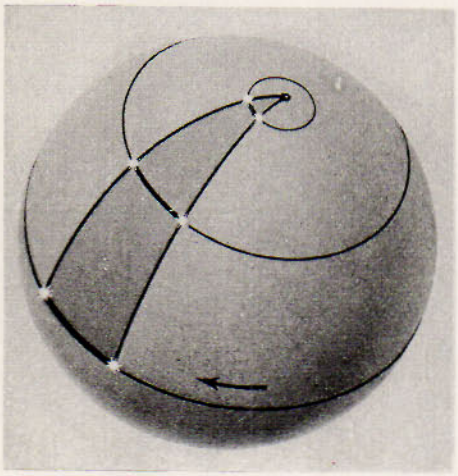


Fig. 118 — LENGTHS OF DIFFERENT PARALLELS OF LATITUDE

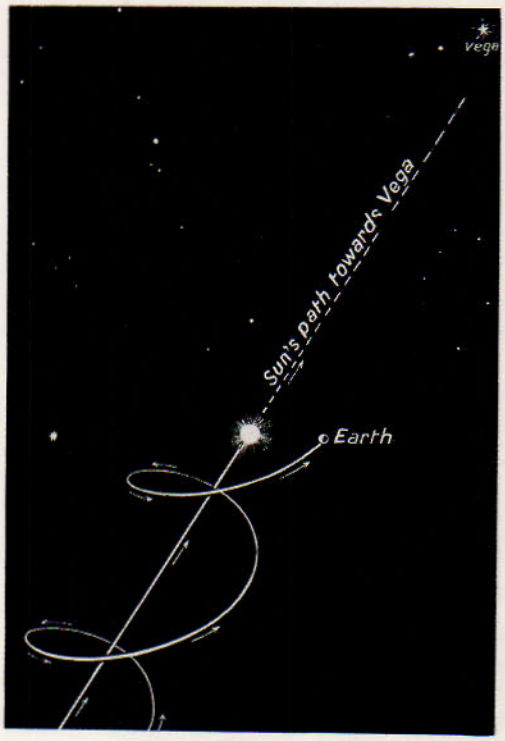


Fig. 119 — EARTH'S TRUE PATH IN SPACE COMPARED WITH THE SURROUNDING STARS

different directions. The Sun can only arrive at the position they occupy at the present moment, which they will have abandoned long before the Sun arrives there. So our Sun will be as isolated as before, unless the star in question happens to be coming to meet it exactly head on, or unless another crosses its path. However, the event is so improbable that it need occasion no anxiety.

These chapters have been mainly concerned with the Earth, its movements and its relation to the Sun, an understanding of which enables us to appreciate the spectacle provided by the stars. Let us now go on to make a more detailed study of these worlds in space, beginning with those which, together with or own, make up the empire of the Sun.

Fig. 120 — THE REGION IN SPACE (THE APEX) FOR WHICH THE SUN IS HEADING

BOOK II

THE EMPIRE OF THE SUN

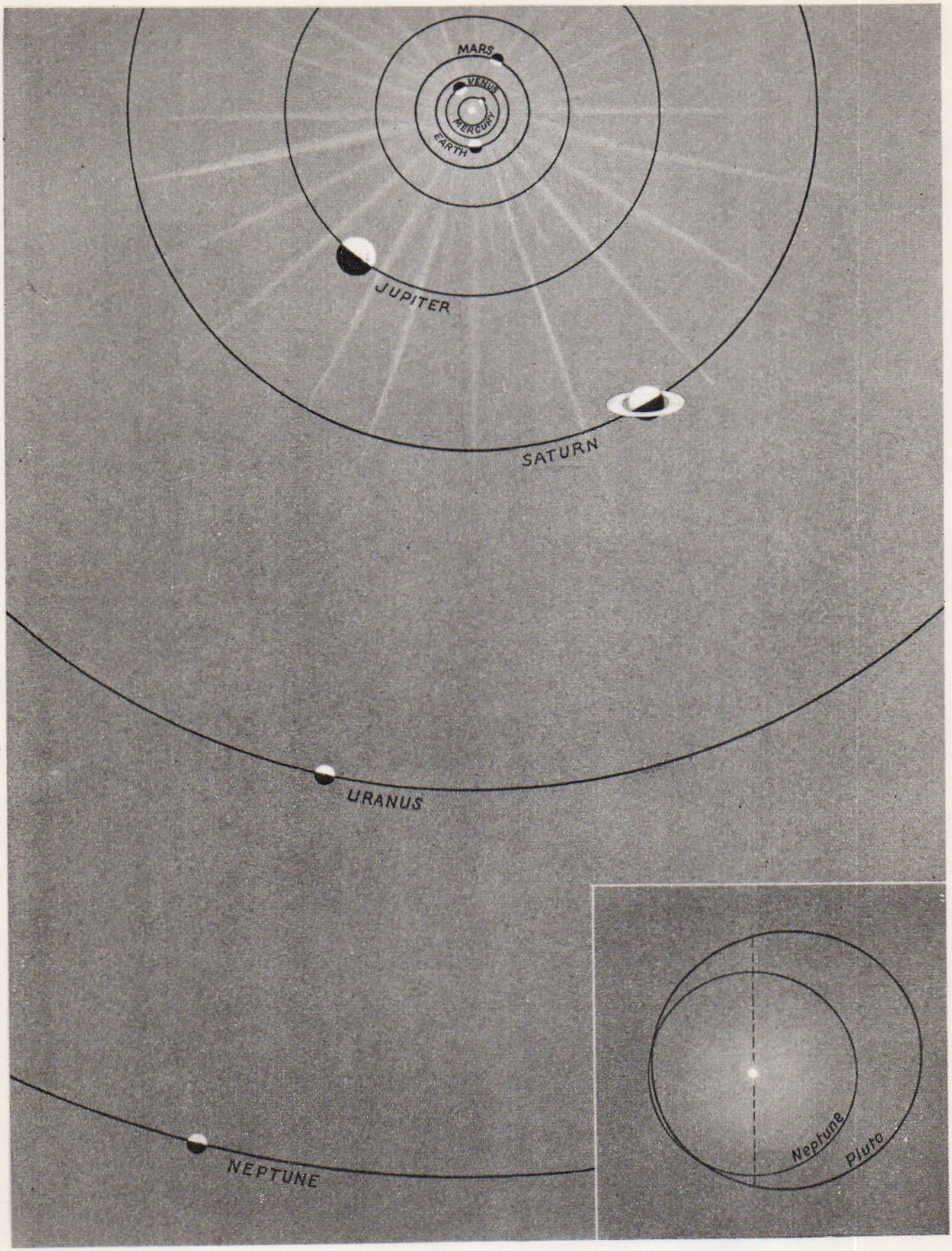


Fig. 121 — CHART OF THE SOLAR SYSTEM

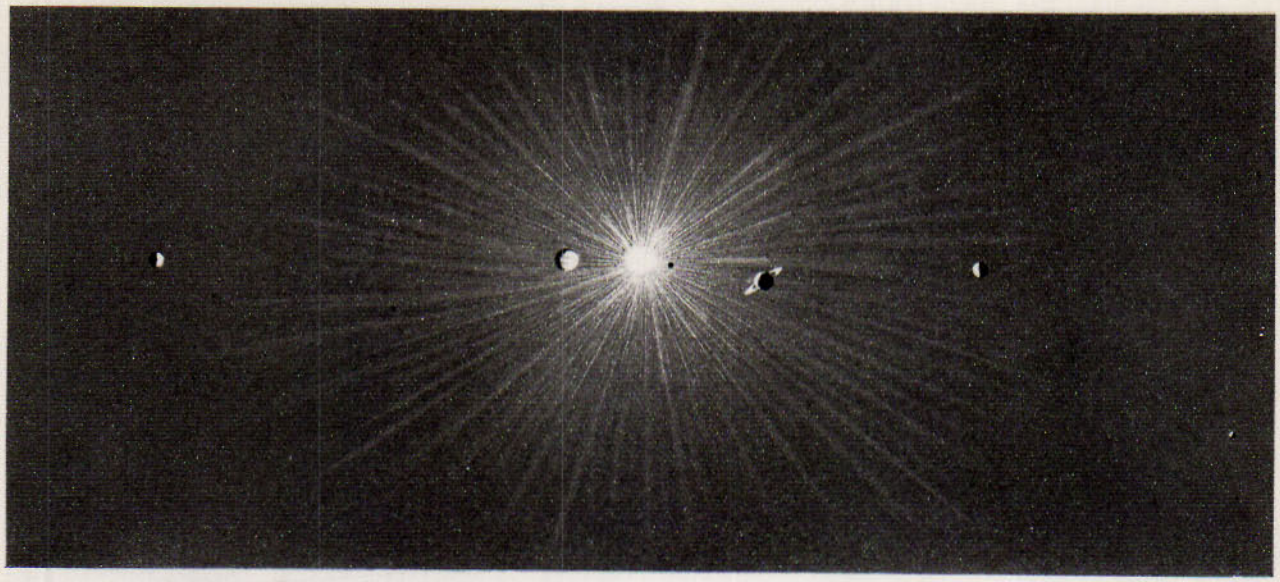


Fig. 122 — THE PLANETS GO ROUND THE SUN IN ALMOST THE SAME PLANE

CHAPTER 3

THE SOLAR SYSTEM

The planets in general. — The general survey of the universe that we made in the preceding chapters showed how small and unimportant the Solar System really is compared with the Galaxy alone. If, however, we examine it in detail from the point of view of the Earth, which is just one of its members, we shall see that the Solar System is very large indeed, so large that, even before we reach the outermost planet, whose orbit may be thought of as the furthest known boundary of the Sun's kingdom, we shall have met several worlds so far away from the Earth that they cannot be seen with the naked eye.

When the first astronomers noticed the 'wandering stars' moving against the background of the fixed stars, and when their immediate successors came to study the exact nature of planetary movements, they could only take into account the nearest and largest of the planets, for these alone are visible to the unaided eye. In order of increasing distance from the Sun, they are: Mercury Venus, Mars, Jupiter and Saturn. For a long time, the ancient astronomers committed certain errors, for it was first thought that Mercury and Venus were not each one body as we know today, but two, appearing before the Sun in the morning sky and after the Sun in the evening. We shall see the reason for this when we come to examine the positions of these two planets in the Solar System, and the way in which they move.

The invention of the telescope was followed by the addition of new planets to those already known.

The first of these new planets, Uranus, was discovered by William Herschel in 1781; it is so far away from the Sun that its discovery doubled the size of the Solar System, whose outer boundary was formed at that time by the orbit of Saturn.

Then from 1801 onwards, astronomers began to discover large numbers of minor planets, or 'asteroids', travelling round the Sun between the orbits of the major planets in a veritable swarm. To this day they have not been properly counted, since fresh discoveries are constantly adding to their number. Minor planets are investigated nowadays mostly by photography and our modern techniques have revealed so many that astronomers have aptly dubbed them 'vermin of the skies'.

The process of extending the limits of the Solar System was continued in the middle of last century when Neptune was discovered at nearly twice the distance of Uranus. The amazing thing was perhaps not so much the discovery itself but the events which led up to it. In this case, it was not made after a long, patient searching of the skies or some happy accident which brought into the astronomer's ken a planet never before seen by human eye: the existence of Neptune was demonstrated by the power of mathematics alone, by the work of two young men who reached the same conclusion unknown to each other. The story is worth retelling.

In an earlier chapter we discussed the perturbations which one body can cause in the movements of another. Soon after the discovery of Uranus, it was realized that the path it was following did not correspond exactly to its theoretical orbit, and it was rightly concluded that Uranus was being perturbed, possibly by Jupiter or Saturn. After a while, however, it became obvious that this hypothesis could not explain the observed discrepancies, and people were forced to the conclusion that they could only be explained on the hypothesis of an unknown planet situated at some distance beyond the orbit of Uranus. It only remained to find this other planet, which is what Adams in England and Le Verrier in France independently undertook to do.

The problem facing them was immensely difficult: they had to analyse all the perturbations to be found in Uranus's motion and from their sense and magnitude deduce the elements of the unknown planet: its distance

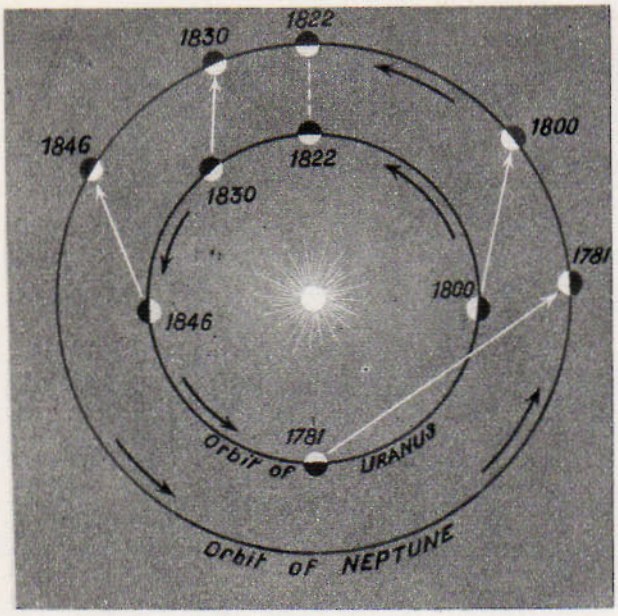


Fig. 123 — Relative positions of Uranus and Neptune from 1781 to 1846, showing how the perturbations occurred

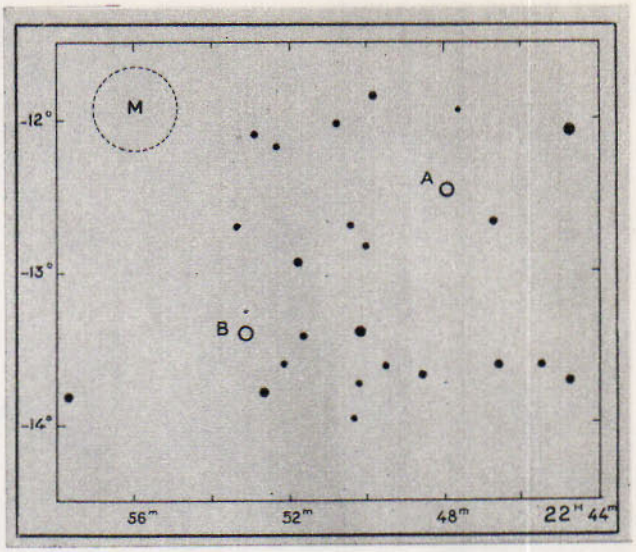


Fig. 124 — Discovery of Neptune. A, the calculated position; B, actual position. The diameter of the Moon (M) is shown on the same scale.

from the Sun, its mass and the position it must occupy in relation to Uranus. They had of course to have some numerical data to use as a starting point, and this was furnished by an empirical law known as Bode's Law (*see below*), which allows us to establish a sort of relationship between the distances of the various planets from the Sun. The approximate distance obtained from this law was narrowed down to its closest limits, and the two men were then able to estimate the planet's mass and probable movement. On 31 August 1846, Le Verrier was able to communicate his results to the Académie des Sciences, Paris, informing them that the new planet lay at a distance thirty times greater than the distance of the Earth from the Sun, and stating its position in the sky. The position given by Le Verrier was 5° from the star δ Capricorni, and in fact, three weeks later, Galle of Berlin Observatory found the planet in almost exactly the predicted position. Meanwhile, Adams, in England, had already reached the same conclusions, and had made known his results to Airy, the then Astronomer Royal. There was some delay in starting a search owing to Airy's scepticism about Adam's results, and when it was finally started by Challis at Cambridge, the necessary sky-charts had to be redrawn, which took up valuable time. Whilst this search was still in progress Galle announced the discovery from Berlin.

There then broke out a most regrettable controversey, in which neither Adams nor Le Verrier took part, about who should be given the honour of the discovery. Posterity has settled the argument by awarding the two men equal fame.

Continuing our exploration of the Solar System, we next come to Pluto. Although the Sun's domain was extended tremendously by the discovery of Neptune, almost the same thing happened again as recently as 1930. For the same reasons that led to the discovery of Neptune, astronomers began to suspect the existence of an even more remote planet beyond the latter, and two or three began to compute its position or search for it telescopically. Unfortunately such a planet would be so far away and therefore so small and faint that the task of finding

Fig. 125 — Neptune's position in the sky at the time of its discovery. It has not yet returned to this point

it among the immense numbers of faint stars in the background and identifying it by its movement would be Herculean. One of the most industrious searchers was the American astronomer Percival Lowell (of Mars fame) and although he himself did not have the success of Le Verrier, his work was continued at the observatory he founded at Flagstaff, Arizona. A systematic photographic search was made by a young astronomer named Clyde Tombaugh and he was rewarded, on 23 January 1930, by the discovery, as a faint spot on a negative, of the planet now called Pluto. Until some new discovery is made, the orbit of Pluto marks the outermost boundary of the Solar System.

Once the elements of a planet's orbit are established, we can determine the path it will follow against the background of the stars. And just as it is possible to predict its future position it is also possible to work out where it has been at any time in the past, before being discovered.

By going through astronomical archives it was found that between its discovery in 1781 and the year 1690, Uranus had been observed more than a dozen times and had been mistaken for a star, and that in the same way, Neptune had been seen three times prior to 1846, on 8 and 10 May 1795 by Lalande, and 25 October 1845 by Lamond. Similarly, Pluto had been photographed at Mount Wilson in 1919, at Yerkes in 1921 and 1927, and also in 1927 at the Royal Observatory, Uccle, Belgium. From this it may be seen that all three planets came very close indeed to discovery on these occasions, but this does not in any way detract from the glory of the people who actually did make the discoveries.

Before we go on to study their exact relationship one to another and the differences between the various planets, let us examine the numerical importance of the Sun's family of worlds. It is usual to group them into the following classes: planets, minor planets (asteroids) and satellites. The planets, as we know them today, are Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, Neptune, and Pluto. They move round the Sun in what are to all intents and purposes concentric orbits, and their mean distances approximately obey a certain law. We may recall that, apart from the Earth, which is the world from which we view them, only five of these planets are visible to the naked eye, Uranus being a border-line case.

This is the simple, classical arrangement of the Solar System as taught in schools, but our ideas must come to embrace a much greater degree of complexity before they can be said to represent the reality. The discovery of numerous asteroids has destroyed this apparently harmonious simplicity: even to show all their orbits at once on a chart of the Solar System would lead to confusion. For the moment, let us say that most of them are contained between the orbits of Mars and Jupiter; some, however, travel further out than this whilst others, because of their

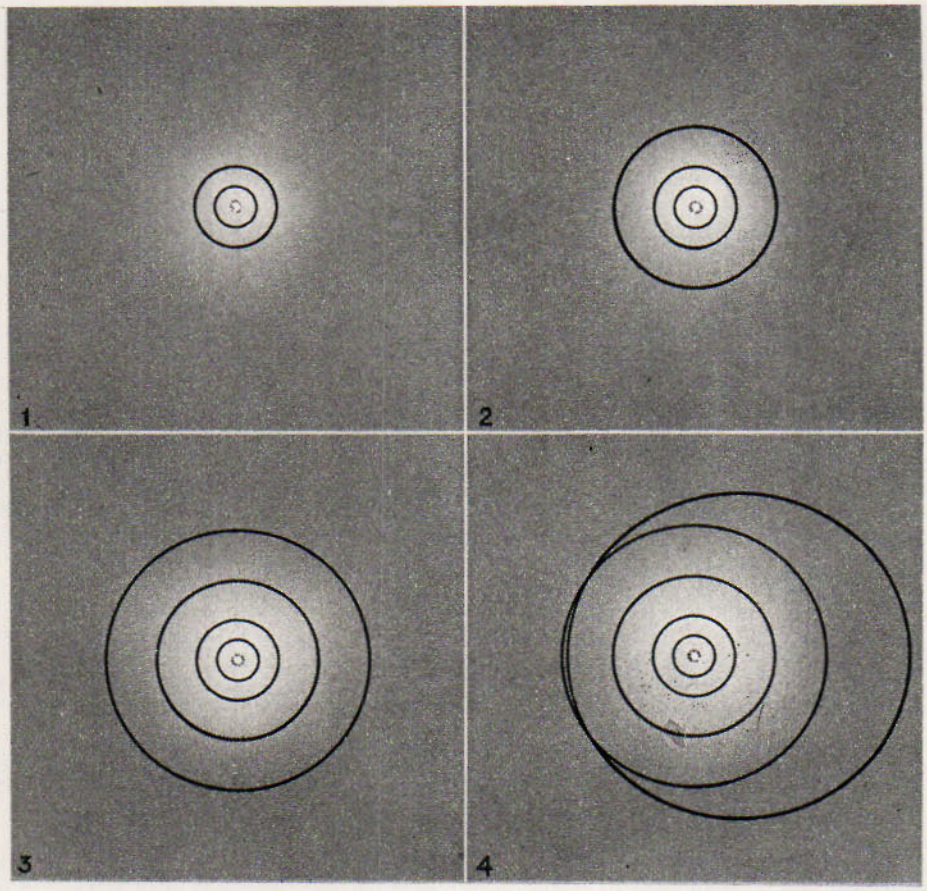


Fig. 126 — SUCCESSIVE STAGES IN THE EXPLORATION OF THE SOLAR SYSTEM. (1) The naked-eye system, ending with Saturn, as known from the most ancient times to 1781. (2) After the discovery of Uranus in 1781. (3) After the discovery of Neptune in 1846. (4) After the discovery of Pluto in 1930. The orbits of the inner planets, out to Mars, are not shown.

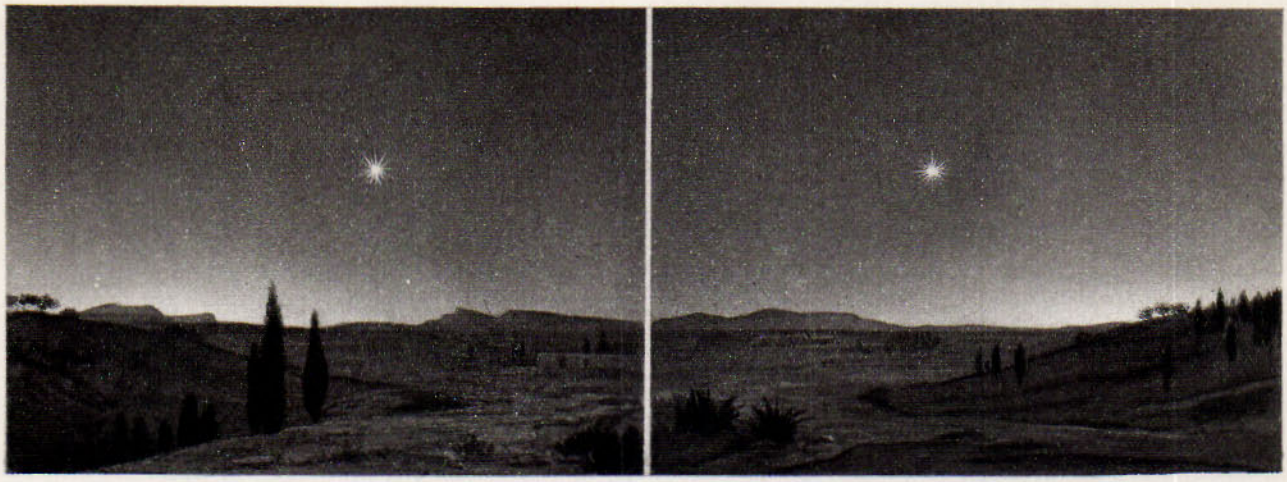


Fig. 127 — MERCURY (OR VENUS) IS SEEN ALTERNATELY AS A MORNING STAR AND AN EVENING STAR

highly eccentric orbits, pass within the orbits of the Earth, Venus and even Mercury.

The number of minor planets is considerable, and indeed, cannot be stated with any accuracy, because every year sees fresh discoveries. If we gave the number known at the moment of writing, it would be out of date before these pages were in print. Up to the present, about 1,500 of these tiny worlds are known for certain, and how many remain to be discovered, it is impossible to say. It is certain, however, that as more systematic searches are carried out with more suitable instruments, the number will continue to grow rapidly. From various considerations, Stroobant, a Belgian astronomer, has calculated that the total number of asteroids can be put between 50,000 and 100,000. On this count we may safely predict that for quite a long time to come, asteroid hunters will continue to be rewarded with success.

The satellites of the major planets make up quite an appreciable part of the Solar System. They are distributed as follows: the Earth has 1, Mars 2, Jupiter 11, Saturn 9, Uranus 5, and Neptune 2. But here again the figures may have to be modified in the future.

As far as our knowledge goes at present, therefore, the Solar System consists of a fairly large number of planets, minor planets, satellites and cosmic debris of various sorts, spread over an area about 6,000,000,000 miles wide.

Nevertheless, we must only take this last figure as a convenient one for reference: all that was said at the beginning of this chapter leads us to bear in mind that we may not yet have reached the limits of the Solar System. It is not unreasonable to suppose that perhaps new ideas leading to fresh research programmes carried out with more powerful instruments, or, as has happened in the past, some completely unforeseen accident, will one day reveal one or more planets beyond the orbit of Pluto.

Reversing this process, we can wonder whether some other planet might not exist in the immediate neighbourhood of the Sun, that is to say, within the orbit of Mercury. At a certain period of the last century, this question was raised very seriously and at one time appeared to have been answered in the affirmative. We shall make a résumé of the facts which inspired this hope, for although no intra-Mercurian planet was ever discovered, the observations which started the whole affair off and which have occasionally been repeated right down to the present day, have never been fully explained.

The original ideas came from the mathematical work of Le Verrier analysing the movement of Mercury: he brought to light a small, secularly increasing difference between the theoretical and observed positions. Le Verrier thought he could explain this discrepancy by perturbations due to one or more small planets situated between Mercury and the Sun.

If something of the sort did exist, it should be visible in the immediate neighbourhood of the Sun before sunrise and after sunset, in the same way as Mercury and Venus. But we have only to consider how difficult it is even to see Mercury regularly against the bright sky after sunset, and we shall realise that a smaller planet much closer to the Sun is going to be practically invisible. However, there are other ways of detecting it. First, during a total eclipse of the Sun, it is possible to search the sky in the Sun's immediate neighbourhood during the few moments of darkness when even faint stars can be made out. Second, as a result of the combination of its movement with ours, the hypothetical planet would at some time or other pass between the Sun and Earth, when it would appear as a tiny black dot moving across the solar disc.

The exploration of the Sun's immediate neighbourhood has often been tried, but without success. On the other hand, there are occasional records of small black bodies passing across the face of the Sun. Although these observations are not in very close agreement, more weight has been attached to the latter sort because of their positive character.

The most remarkable of these observations was made by an amateur astronomer, Dr Lescarbault, on 28 March 1859.

As this phenomenon seemed to justify the theoretical conclusions he had arrived at, Le Verrier set about analysing all the observations of this sort that he could find. Of the fifty or so cases he collected, the majority were doubtful or insufficiently documented. He was left with six, from 1802, 1819, 1839, 1849, 1850 and 1861, which he believed he could connect with a single body. Working with elements derived from assuming this hypothesis, Le Verrier calculated several orbits which satisfied the required conditions; nevertheless there was one which seemed preferable to the rest and which demanded that the planet, a little prematurely christened 'Vulcan', should have a revolution period of 33 days, in an orbit strongly inclined to the ecliptic. From this orbit and period, it was announced that there would be a transit across the Sun on 22 March 1877. The astronomers of the world waited in vain that day.

Since then there has been nothing to confirm the existence of the hypothetical planet, but neither has there

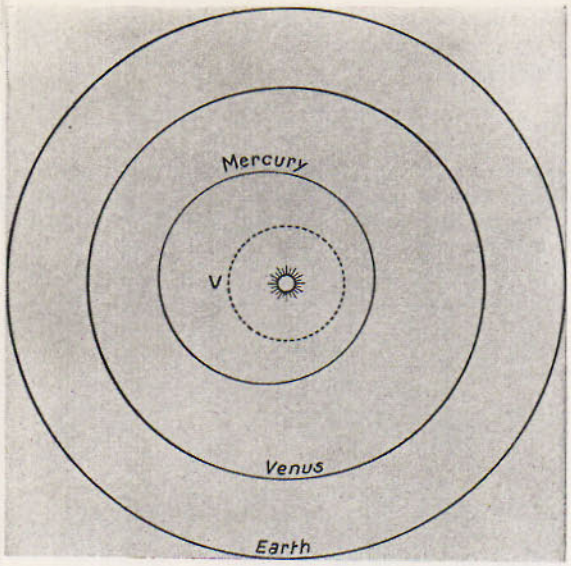


Fig. 128 — THE ORBIT OF THE HYPOTHETICAL PLANET VULCAN, AFTER LE VERRIER

been any explanation of reports concerning very small bodies passing in front of the Sun's disc, which still come in from time to time.

First, people have blamed those small celestial bodies which, wandering round the Solar System in all directions, sometimes encounter the Earth and produce the brilliant spectacle we call 'bolides'. There is no proof of this hypothesis, and there is equally good evidence in favour of another one, following recent discoveries of unusual minor planets. Some of these have such eccentric orbits that they approach nearer the Sun than the inner planets. Consequently, it is certain that at least some minor planets can pass between the Sun and Earth, but here we are entering on a field of knowledge which has not yet been properly investigated: many minor planets undoubtedly remain to be discovered, and many of those that are known have not been observed long enough for their orbits to be determined with sufficient accuracy. In the present state of our knowledge, therefore, it would be very difficult to attribute the recorded observations to any particular asteroid or asteroids. Nevertheless, this is very probably the answer to the problem.

In any case, it seems fairly well established that there is no body of any size in the immediate vicinity of the Sun, and we may conclude that the description of the Solar System advanced in the next section is substantially correct in this respect.

The distribution of the planets in the Solar System. — The orbits of the planets are arranged in order of increasing distance from the Sun (*see* Fig. 121). This diagram, necessarily on a very small scale, can give only a general view of the arrangement, and consists merely of concentric circles drawn round the Sun, which forms the centre of the system, each circle having a radius which corresponds to the mean distance of the respective planet. This brings out the relative dimensions of the various orbits. However, we cannot be content with this very elementary information.

In order to understand the exact relationships and the movements involved, we must describe in detail the main characteristics of the orbits, and their orientation with respect to the plane of the ecliptic, with which each makes a different angle. The latter, indeed, is an aspect which a plan view of the Solar System does not bring out, for, on a plan view, the orbits must be represented as if they were all in the same plane. It is important to emphasise this because certain orbits appear to cross when seen in plan, which may lead to confusion. In fact, the eccentricity of a particular orbit may mean that its perihelion may be nearer the Sun, or its aphelion further away in relation to another orbit, beyond or before which it is situated at mean distance. On a drawing, therefore, the two will seem to cross like roads on a map, whereas in fact they are superimposed on account of the angle between them. At the point where one orbit appears to cross the other, one is above the other, as in the case of a bridge (Fig. 129).

In comparing the arrangement of the various orbits, it is necessary to consider the direction in which the two orbital planes intersect (we have already met this in considering the way in which the Earth's equator intersects the ecliptic). In the Solar System, orbital planes are likewise referred to this fundamental plane of the ecliptic. The orientation of an orbit is defined by the longitude of its perihelion — thus fixing the direction of the major axis — and by the nodes, or intersection points of the two planes. A planet moving along its orbit is said to pass through the ascending node when it cuts the ecliptic from south to north on the celestial sphere, and through the descending node when in the opposite direction.

All the planets revolve in the same direction, called direct, which for an imaginary observer placed at the north pole of the ecliptic is counter-clockwise. The same

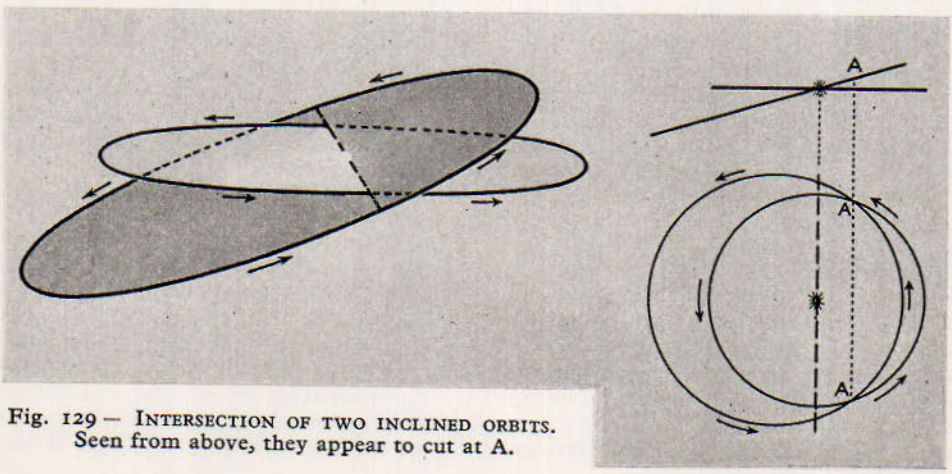


Fig. 129 — INTERSECTION OF TWO INCLINED ORBITS. Seen from above, they appear to cut at A.

applies to the satellites of the major planets, with certain exceptions when the movement is said to be retrograde.

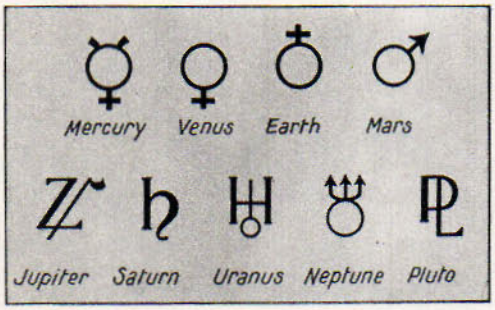
The table below sums up the principal elements of each planet.

It will be seen that this table deals only with the planets themselves, the asteroids being too numerous to include. More details about these minor planets, which are situated for the most part between the orbits of Mars and Jupiter, and the various problems they raise, will be given in the appropriate chapter.

NAME	SEMI-MAJOR AXIS OR ORBIT IN A. U.	MEAN DIST. IN MILLIONS OF MILES.	SIDEREAL REV. PERIOD.	ECCENTRICITY	INCLINATION
MERCURY	0.387	36.0	87.969d	0.205	7°00
VENUS	0.723	67.2	224.7d	0.006	3°23′
EARTH	1.000	92.9	1y 0.006d	0.016	
MARS	1.523	141.5	1y 321.729d	0.093	1°51′
JUPITER	5.202	483.3	11y 314.839d	0.048	1°18′
SATURN	9.554	886.1	29y 166.98d	0.055	2°29′
URANUS	19.218	1783	84y 7.45d	0.046	0°46′
NEPTUNE	30.109	2793	164y 280.3d	0.008	1°46′
PLUTO	39.60	3666	249y 76d	0.246	17°7′

As for our own world, whose rôle of moving observatory we shall use to explain the behaviour of the planets as seen in the sky, we have given it its rightful place among its fellow planets. The Earth has but one feature to distinguish it from the rest: it is our home.

Finally, the following symbols are used to designate the various members of the Solar System:



Bode's law. — We have already mentioned the fact that the distribution of the planets by increasing distances from the Sun seems to follow some regular pattern.

When we compare these distances, we realise that they more or less conform to a 'law' which was recognised by the German, Wolf, as early as 1741 taken up again by his compatriot Titius in 1772 and finally formulated mathematically by Bode, in 1778 This law is usually named after the latter, and can be established as follows:

Take a series of numbers beginning 0, 3, etc; if each figure is double the one preceding it (except in the case of the first two) we then have the following series:

0 3 6 12 24 48 96 192 ... ...

If we now add 4 to each of these numbers, we have

4 7 10 16 28 52 100 196 ... ...

Now the relationship between these numbers is more or less the same as the one between the distances of the various planets from the Sun, taking the Earth's distance as 10. If we multiply the first numbers in the foregoing table by 10 (for by common usage, the Earth's mean distance constitutes one Astronomical Unit), the distances of the planets known in Bode's time correspond quite closely:

Bode's series:

4 7 10 16 28 52 100

The planets:

3.9 (Mercury) 7.2 (Venus) 10 (Earth) 15.2 (Mars) 52.0 (Jupiter) 95.5 (Saturn).

This satisfactory agreement seems to be supported by the discovery of Uranus lying at a distance of 192 A. U., which is very close to the 196 of Bode's series. There was only one case of disagreement — or rather, a disturbing gap: the number 28, as the reader will already have noticed, did not correspond to any known planet. This naturally led to the conclusion that there might be some unknown object in this position which had so far escaped detection, and a committee of astronomers was formed with the aim of searching for it along the Zodiac. Their systematic work had proved fruitless when, quite by chance, on 1 January 1801, Piazzi, working at Palermo, noticed a small star-like object that he had never seen before. The next day he observed that it had moved, and this movement continued day after day. For a short while it was thought to be a very small comet (as in the case of Uranus twenty years earlier) but before long it was recognised as a new planet, to which Piazzi gave the name Ceres. When the elements of its orbit were calculated, its distance of 27.6 A. U. filled the gap in Bode's series.

Later, Bode's law, as we have seen, was useful as a starting point in Le Verrier's work on Neptune, since it gave an approximation to the spacing of the planets in the Solar System. His first numerical datum was the figure 388, which represented the position of the unknown planet if it were valid to continue the series. However, it must be added that as soon as its existence was confirmed and once the elements of its orbit had been calculated from direct observation, this figure had to be modified considerably. Instead of the empirical distance of 388, which to satisfy theory had been reduced to 330, the distance was finally reduced even more drastically to 301. This considerable difference shows that although Bode's law played a very important part in the discovery of Neptune, the discovery itself put the law in doubt! With the discovery of Pluto, the law fails even more seriously, since the planet is situated at a mean distance of 396, which ought to be 672 if the original series is continued.

Nevertheless, it is still true to say that however empirical it may be, Bode's law has played a most important part in the exploration of the Solar System.

Comparative dimensions of the planets. — The planets are not only situated at very different distances from the Sun, but they also vary considerably in size,

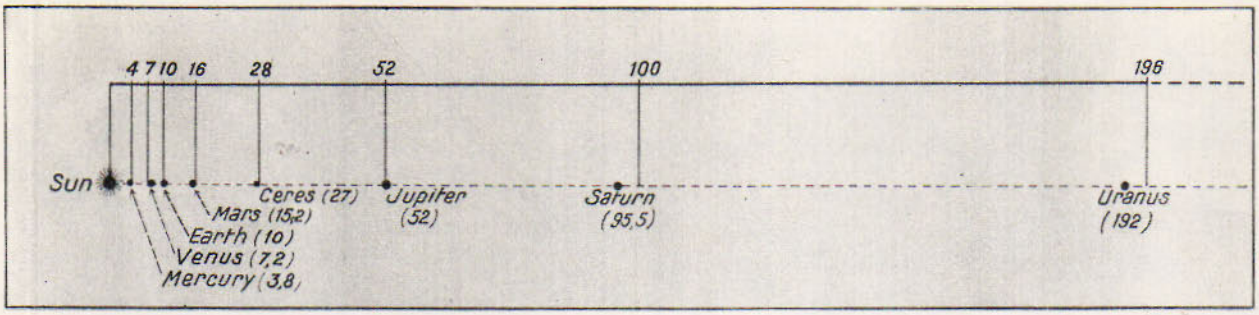


Fig. 130 — BODE'S LAW

Compared with the Earth, some are enormous, some are about the same size and some are quite tiny. Before discussing them in detail, it is worthwhile pointing out right from the start the most essential facts about the planetary system, which will make it easier to draw comparisons. The accompanying table summarises the main facts.

NAME	DIAMETER		VOLUME	MASS	MEAN DENSITY	
	EARTH = 1	MILES	EARTH = 1		EARTH = 1	WATER = 1
MERCURY	0.37	3,100	0.05	0.06	1.1	6.2
VENUS	0.97	7,700	0.90	0.82	0.91	5.0
EARTH	1	7,927	1	1	1	5.52
MARS	0.54	4,200	0.1	0.11	0.69	3.81
JUPITER	11.1	88,700	1295	318.36	0.25	1.36
SATURN	9.4	75,100	745	95.22	0.13	0.72
URANUS	4.0	30,900	63	14.58	0.23	1.27
NEPTUNE	4.3	33,000	78	17.26	0.22	1.20

For the same reasons as in the table of the orbital elements, the asteroids have been omitted here, and they will form the subject of some special remarks later on. Pluto also is absent, for as yet we have very little definite information about the dimensions of its globe.

It will be seen at first glance that there are important differences between the planets and the Earth in many respects; the world we live on seems vast to us but in truth it cuts a very poor figure beside such giants as Jupiter and Saturn. In fact we can distinguish two groups of planets: the giants, which are composed of materials with a very low density, and the dwarfs, which are almost the same size and have roughly the same density as the Earth. The first group, beginning with Jupiter, comprises the most distant planets from the Sun; Pluto must be omitted for its diameter is very small and is not known with any great certainty, whilst its mass, which must be comparatively large, indicates the highest known density in the Solar System. The second group includes all the interior planets, that is to say, as far as and including Mars. Between these two groups lies a swarm of minor planets, their exact nature not yet decided, but which are all extremely small. The largest of the asteroids is Ceres, the first to be discovered, which has a diameter of 400 to 500 miles; Pallas cannot be much more than 300 miles in diameter, Vesta 250 miles and Juno about 130 miles. These are the only asteroids whose diameter it has been possible to measure with any accuracy: all the others must be deduced as best we may from photometric estimates. It has been found in this way that some cannot exceed more than 3 or 4 miles across, and there are others which must be even smaller.

We find the same differences between the satellites as between the planets. As we shall see, the satellites of Mars are scarcely larger than small asteroids, and this applies to certain members of the systems of Jupiter and Saturn, where we shall also find satellites as large as or even larger than Mercury itself.

It should not seem surprising that so many of these dimensions are given in the form of approximations. When we study how they are measured, the immense difficulty of the task will become apparent. We must remember the enormous distances involved: a large body at a very great distance appears as no more than a tiny

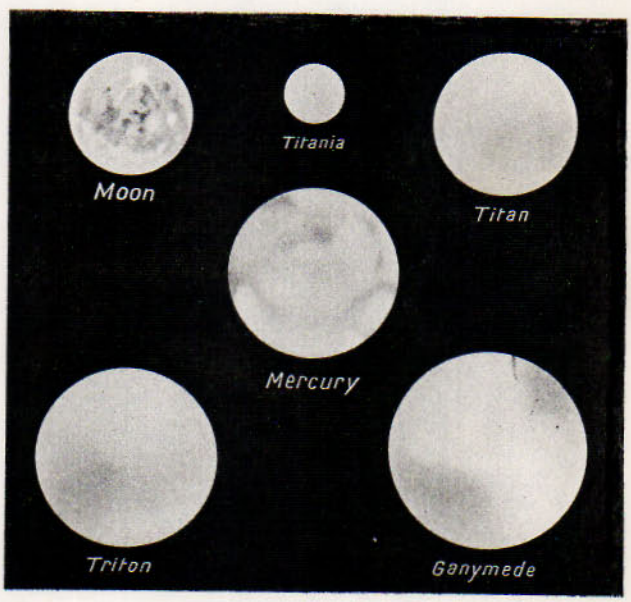


Fig. 131 — THE LARGEST SATELLITE OF EACH PLANET COMPARED WITH THE MOON AND WITH MERCURY. Both the satellites of Mars are too small to be worth showing on this scale.

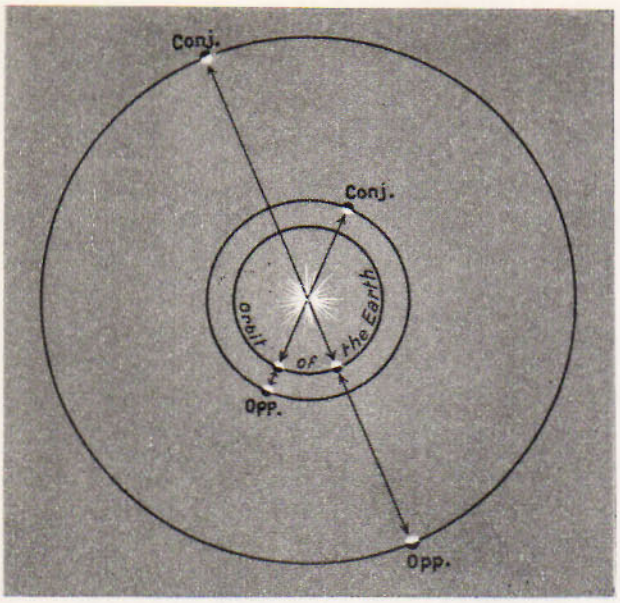


Fig. 132 — THE FURTHER FROM EARTH THE PLANET, THE LESS ITS RELATIVE DISTANCE VARIES BETWEEN OPPOSITION AND CONJUNCTION

dot in our instruments, and this of course makes it all the harder to measure. This is only one of a large number of factors which can vitiate the accuracy of measurements, introducing a noticeable error into the results.

In spite of these difficulties, however, our figures on the whole are sufficiently accurate for us to have a good idea of the real sizes of the other planets, and the conclusions which may be drawn from them allow us to make striking comparisons with our own world. In fact, the expression 'other worlds' has emerged from the realm of speculation where it was born and, thanks to reliable physical data about the heavenly bodies, has become a sort of living reality.

We cannot, however, deny that much still remains to be discovered. However powerful the instruments at our command, they are hampered by factors which are almost or completely insurmountable at the present time. These factors are due to circumstances both within and outside our own planet, and we are thus led to consider the conditions under which we, stationed on the Earth's surface, are forced to observe the planets.

The appearance of the planets and the Earth's rotation. — The observation of planets is governed largely by their distance and situation relative to the Earth and the Sun. All the members of the Solar System, including the Earth, are moving in orbits round the Sun; this means that the distance between any two planets is constantly varying.

The first consequence of this is to produce an alternate increase and decrease in the size of a planetary disc which of course attains its maximum diameter when nearest the Earth, that is to say when the two bodies are running along the parts of their orbits which are closest together. When the two planets are on opposite sides of their orbits, the situation is reversed and they are at maximum distance. The difference between these two extremes is more pronounced for neighbouring planets than for extremely distant ones (Fig. 132) and their range of apparent diameters is greater.

Taking them all at their mean distances, the apparent diameters of the planets are not so disproportionate as their real dimensions; but for observational purposes, the apparent diameters are more important, since at certain times the nearer planets may become too small to observe. This is shown in Fig. 133.

As will be realised, these often large differences are not the only factor which we discover through observation. The position of a planet with relation to the Sun and Earth determines what phase-effect it will show when

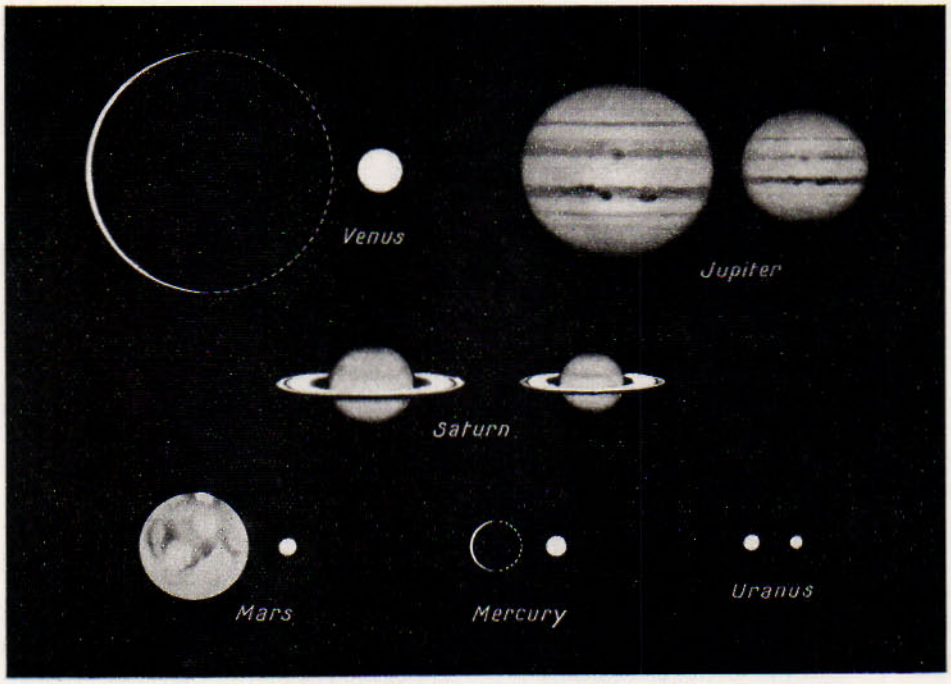


Fig. 133 — VARIATIONS IN THE APPARENT DIAMETERS OF THE PLANETS DUE TO VARIATIONS IN THEIR DISTANCES FROM THE EARTH

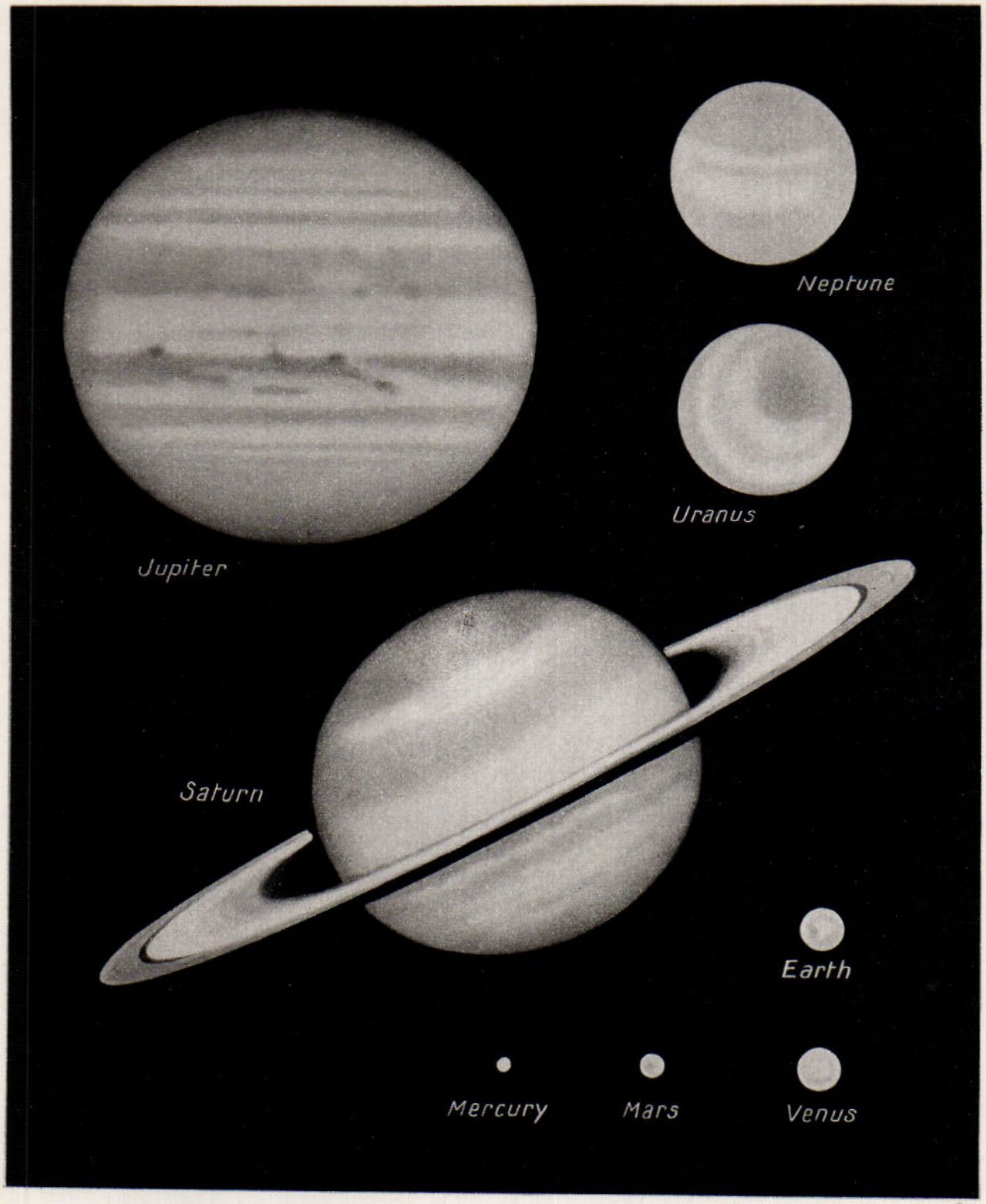


Fig. 134 — RELATIVE SIZES OF MEMBERS OF THE SOLAR SYSTEM. (Pluto is not shown.)

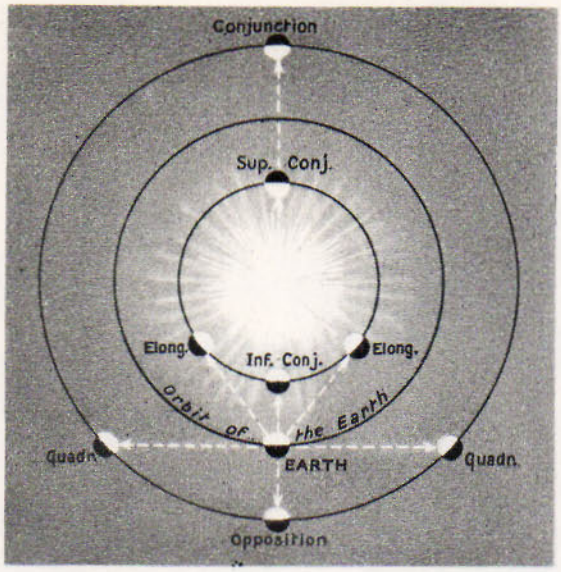


Fig. 135 — PLANETARY MOVEMENTS: CONJUNCTION, ELONGATION, QUADRATURE, OPPOSITION

seen from the Earth, for the Sun shines on only one half of a spherical body and it may well be that all or most of the unilluminated portion is turned in our direction, as frequently happens in the case of the Moon.

We shall deal with this question by simplifying it to some extent, that is, by considering two cases typical of the rest. For this purpose the planets may be divided into two groups: those nearer the Sun than is the Earth are called inferior planets — i.e. Mercury and Venus — whilst the remainder are all outside our orbit and are called superior planets.

We shall consider all the orbits as concentric circles, for here we are emphasising the angles that the planets make with the Sun. The explanation can be followed quite easily from Fig. 135 which shows all the features to be considered in the above cases. When a superior planet is at its nearest to the Earth, the two bodies are both on the straight line joining them to the Sun. Thus a body which is opposite the latter with respect to the Earth is said to be in opposition, and appears as a fully-illuminated disc, since the whole sunward side of the planet is also turned towards the Earth. It is at opposition that a body reaches its maximum apparent diameter. When a planet is diametrically opposed to this position, that is to say when the Sun is directly between it and the Earth, it is said to be in conjunction; if the Sun's dazzling brilliance were not in the way we should see the planet as a fully illuminated disc, but reduced to its minimum diameter. To go from one extreme to the other, it must pass through all the intervening stages and as seen from the Earth, it makes a continuously varying angle with the Sun. During this time, the planetary globe, the illuminated half of which always faces the Sun, is presented obliquely to our view and shows a phase-effect dependent on the angle between the three bodies. This phase is at maximum when the planet is at quadrature, that is at right angles to the Sun.

The apparent movements of the inferior planets — Mercury and Venus — are different because their orbits are completely contained within that of the Earth, and are seen in perspective according to their inclination. On account of this, the movement of Mercury and Venus appears as a sort of oscillation from one side of the Sun to the other. The maximum distance reached to east or west is called elongation (this distance is much greater for Venus than for Mercury because of the former's larger orbital radius); at inferior conjunction they are between the Sun and Earth, whilst at superior conjunction they are on the opposite side of the Sun. At all events, these inferior planets never depart far from the vicinity of the Sun and can be seen to advantage only in the twilight after sunset or before dawn, although Venus is so bright that at times it can be found in full daylight with the naked eye, if one knows exactly where to look.

In the course of their movements around the Sun which illuminates them, they display a series of phases in every way comparable with those of the Moon. This is shown in Fig. 136. Unlike the other planets Mercury and Venus, which pass through inferior conjunction when nearest the Earth, are then most unfavourably placed for observation since at best they appear only as razor-thin crescents of light, the dark half of their globes being almost entirely turned in our direction. And as the diagram also demonstrates, the more of the planet that comes into view as the illuminated portion begins to turn towards us, the smaller the apparent diameter becomes with the planet's constant retreat from us round its orbit.

In this very general explanation, we have only considered the relationships determined by the possible respective situations of the various planets. For example, when we say that a planet is in opposition, this means that, measured on the celestial sphere, the angular distance separating it from the Sun is 180° (Fig. 137), but the planet is not necessarily at the Sun's antipodal point.

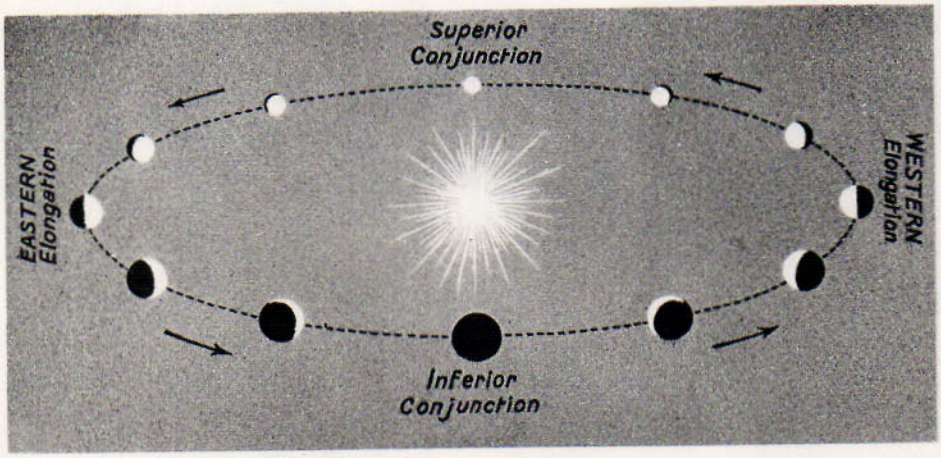


Fig. 136 — EXPLANATION OF THE PHASES OF MERCURY AND VENUS

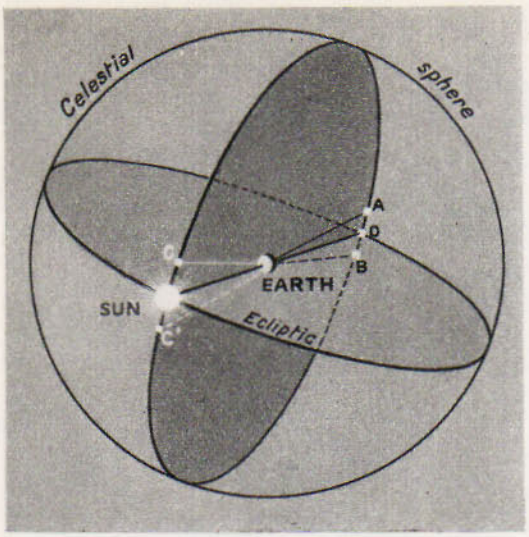


Fig. 137 — INCLINATION OF ORBITS. At opposition a planet may be seen on the celestial sphere at O, A or B; at conjunction, at C or C′.

We must remember that planetary orbits are inclined to the ecliptic by various amounts, and this means, as Fig. 139 shows, that a planet can at this time pass either above or below the antipodal point, which is exactly 180° round the ecliptic from the Sun. For the three bodies to coincide they must be situated on a straight line and this condition is fulfilled when they are at one of the nodes (the points where the two orbits intersect). On this occasion, Venus or Mercury may be observed as a black dot crossing the face of the Sun. Later we shall make a more detailed examination of this interesting and rare phenomenon, for it gives rise to unique observational problems, as indeed do normal inferior conjunctions occurring above or below the Sun.

If we compare various planets, the foregoing account will show that generally speaking, Mercury and Venus are less favourably placed for observation than planets situated beyond the orbit of the Earth, since the latter do not come out until the Sun has ceased to illuminate the atmosphere with its full brilliance: in the evening we have to wait until the Sun has sunk far enough beneath the horizon to observe them while in the morning they become visible an appreciable time before sunrise. This means that since Mercury and Venus are always somewhere in the vicinity of the Sun, that is to say, following it or preceding it in the sky, they can be observed only when near the horizon and in many cases when still in the glow of the twilight. In contrast to this a superior planet can be seen under the most favourable conditions possible: at opposition it is fairly high up in a dark sky, due south at midnight. At eastern quadrature, the planet is in this position about sunset and at conjunction it will be behind the Sun. Then it begins to appear in the morning sky, and at western quadrature it will be due south at sunrise. As the weeks go by, the planet approaches opposition again, and so the cycle continues. We may here recall the remarks made concerning the yearly revolution of the celestial sphere: the planets, like the stars, take part in the daily movement of the sky caused by the Earth's movement round the Sun, but their real motion must be superimposed on this, since it brings about a change in their position relative to the Earth, as we have just explained. So although we see them rise and set every day with the stars (with which they may easily be confused at first sight) in fact they gradually travel completely round the celestial sphere. For each planet, the time taken to do this is theoretically equal to its revolution period, but since the Earth itself is moving, the planet returns to a given position with respect to the latter after a period determined by the interaction of the two movements.

When a planet is at quadrature, it shows a phase-effect, like the Moon, but this is more pronounced for a near planet than for a very distant one. The angle A

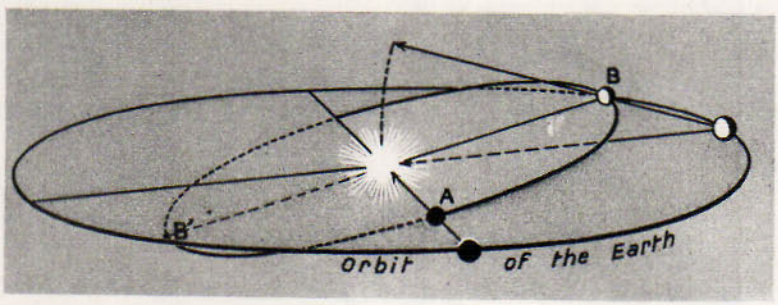


Fig. 138 — INCLINATION OF ORBITS. At inferior conjunction Mercury or Venus may be seen passing in front of the Sun at A, or above it at B, or below it at B′.

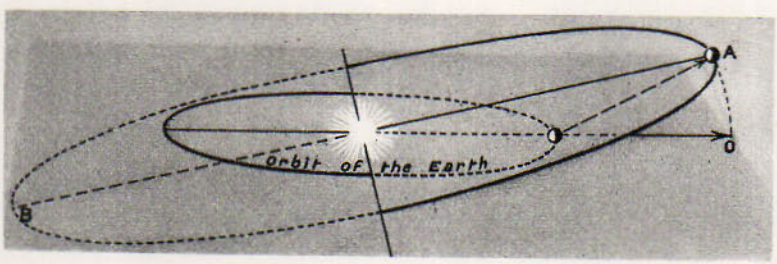


Fig. 139 — INCLINATION OF ORBITS. At opposition, a superior planet at A or B is situated slightly away from the point which is 180° from the Sun.

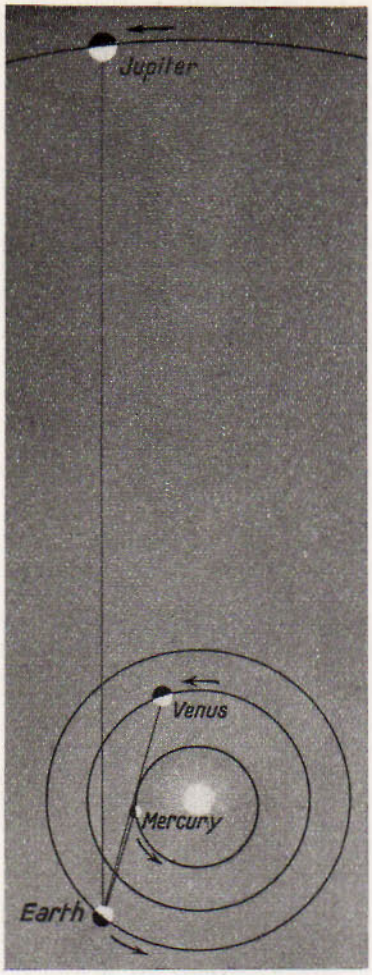


Fig. 140 — RELATIVE POSITIONS OF TWO PLANETS WHEN THEY ARE IN CONJUNCTION AS SEEN FROM THE EARTH

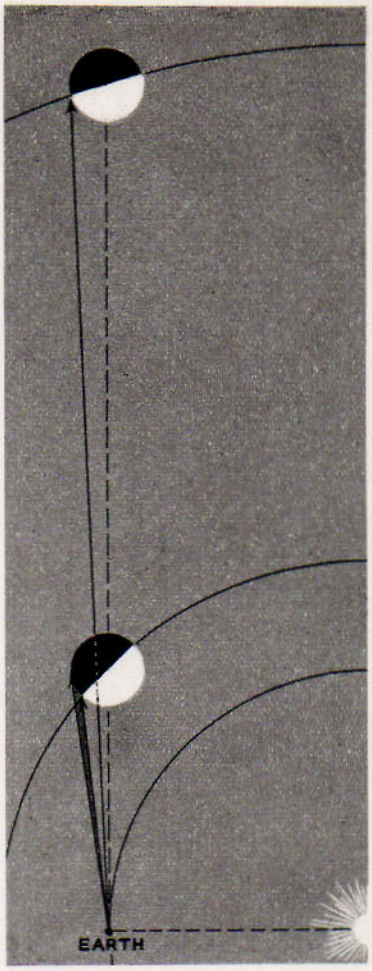


Fig. 141 — PHASE EFFECT ON A SUPERIOR PLANET, ACCORDING TO DISTANCE

(Fig. 140) between the direction of illumination and our line of sight is larger in the first case than in the second; and if the planet in question is one of the outer worlds of the Solar System, the effect becomes quite negligible. Thus, while for Mars, which is relatively close to the Earth, the phase-effect is quite appreciable, Jupiter at four times the distance in the same circumstances shows merely a slight darkening along one edge of the disc, the result of the decrease in illumination at the terminator. The same effect can be observed on Saturn, but is far less prominent. In all these cases, the effect brings out the spherical nature of these worlds. The variations in the direction of illumination compared with our line of sight are also important in connection with the satellites of the various planets. In particular we shall make a study of Jupiter's moons, for while being easily visible they also provide a very good way of demonstrating the mechanism of eclipses.

The apparent movements of the planets. — The foregoing section described the appearance of the planets as seen from the Earth. Now we shall examine the way

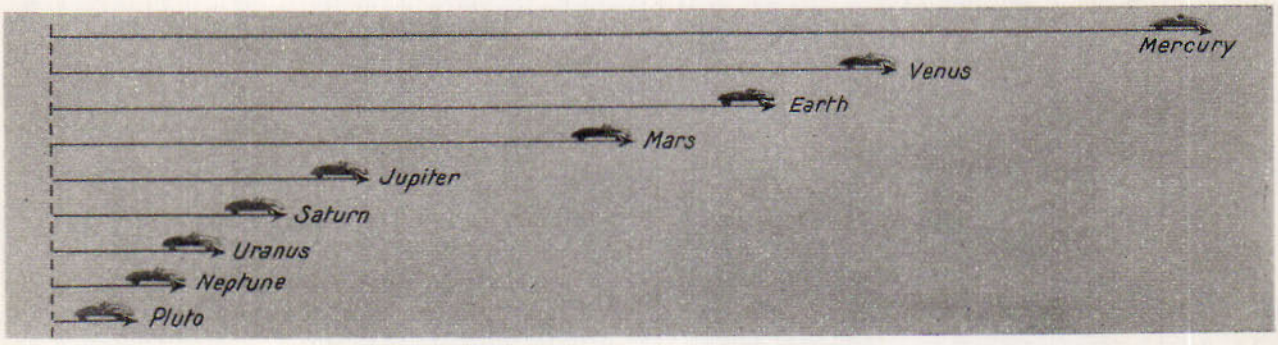


Fig. 142 — THE RELATIVE ORBITAL VELOCITIES OF THE PLANETS

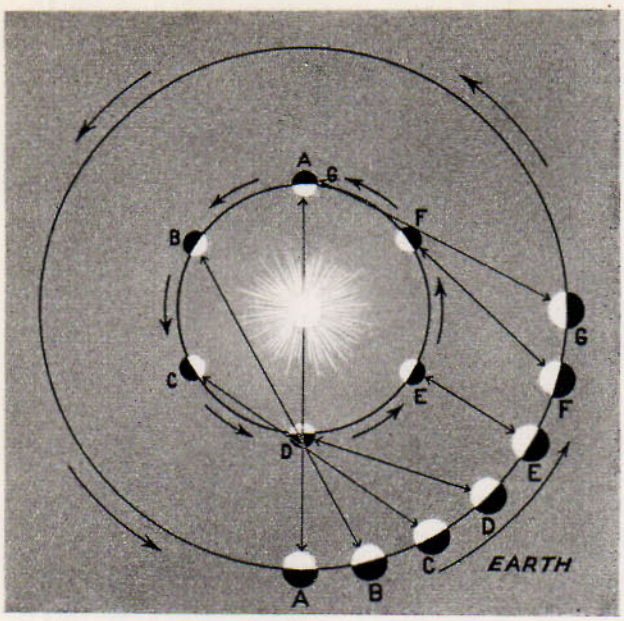


Fig. 143 — COMBINATION OF THE MOVEMENTS OF THE EARTH AND OF AN INFERIOR PLANET

their various movements appear from our moving observation post, the Earth. For the sake of simplicity, we shall again assume that the planets move uniformly in circles round the Sun, for to take into account the slight eccentricity of most orbits and their irregularities would merely obscure the main point we are trying to make.

As we have already seen, all the apparent displacements result from a combination of our own movement with the real movement of each planet, which may be faster or slower than our own. Instead of dealing with each member of the Solar System individually, we shall once again take two typical cases — an inferior and a superior planet — to represent them all.

Once the general mechanism is understood, a little thought will enable us to see how it works in each case, depending on the planet's speed, distance, etc.

For our first example, let us consider a planet lying between us and the Sun: either Mercury or Venus. We have already shown that they are observed first on one side of the Sun, then on the other, each passing through its cycle of phases. The planet's apparent speed will vary according to the direction in which it is travelling with respect to the Earth. We can follow its movement in Fig. 143, starting at superior conjunction. As the planet leaves superior conjunction and moves towards elongation, the angle between it and the Sun increases and its course curves round towards the Earth. Seen in perspective, this movement becomes less and less pronounced, and there comes a time when the planet is moving directly in our line of sight and is apparently stationary.

From now on the planet will seem to approach the Sun again, its speed increasing, and it will pass between the Sun and Earth at inferior conjunction. Then it will emerge at the other side of the Sun, go through the same apparent movements but in reverse order until it finally returns to superior conjunction.

We must emphasise that this must not be confused with the planet's apparent movement with respect to the stars, for it must be remembered that the Earth is moving. On the one hand, the Earth's movement must be taken into account as well as that of the planet to determine the interval after which they return to the same relative positions. On the other hand the Earth's movement will affect the position in which the planet appears against the celestial sphere. We may compare this with the way in which the Sun seems to make a complete journey round the ecliptic in the course of a year; it is the same for a planet except that the latter moves in relation to the Sun. And as a result of all these movements combined with the planet's inclination to the plane of the ecliptic, its apparent course against the background of the stars will be in the form of a series of loops, with stationary periods and retrograde movements.

This explanation will immediately enable us to understand the behaviour of Mercury and Venus. These considerations concerning their apparent movement against the background of stars are more theoretical than practical in importance since the two planets are visible only in or near the twilight glow of sunset and dawn, and cannot be seen in the midnight sky in company with stars which may be used as reference points. Therefore it is their movement with respect to the Sun on which their phases depend and which dominates our attention, whereas with the outer planets, we can measure their movement against the background of the constellations.

Here again, the mechanism we have to deal with is the same as the one above, but in this case it is the Earth which is the fast-moving planet. Fig. 144 shows the same combination of movements, but we experience them differently. We catch up with a planet, overtake it and continue on our way round the Sun until we overtake it once more. During this time, the planet will have appeared to go through the stages defined above — opposition, quadrature and conjunction. The combination of the two movements — that of the Earth and the real movement

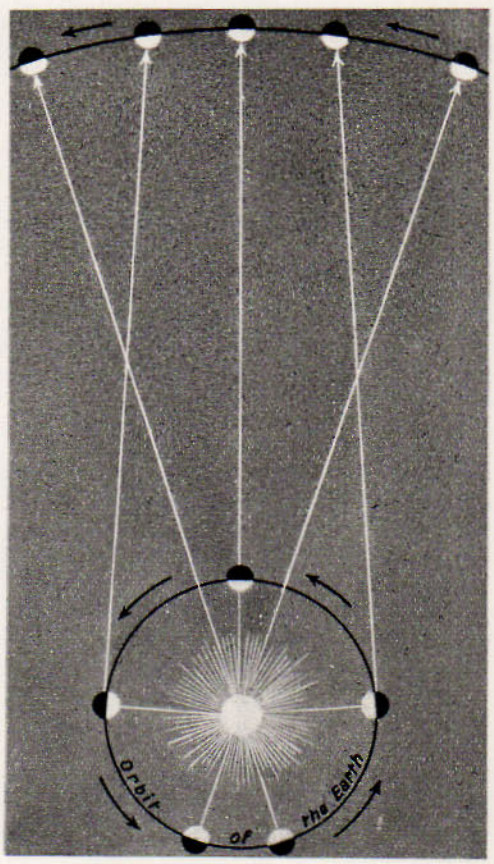


Fig. 144 — COMBINATION OF THE MOVEMENTS OF THE EARTH AND OF A SUPERIOR PLANET

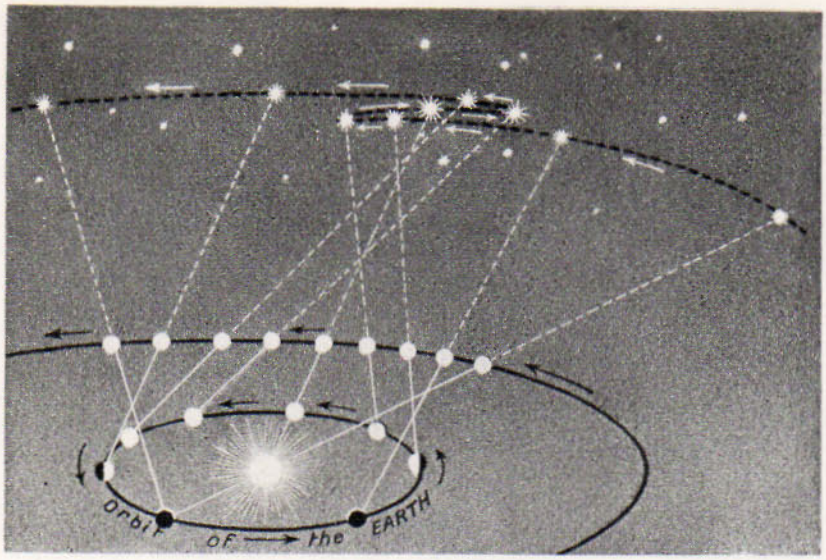


Fig. 145 — COMBINATION OF THE MOVEMENTS OF THE EARTH AND OF A SUPERIOR PLANET, SHOWING STATIONARY POINT AND RETROGRADATION

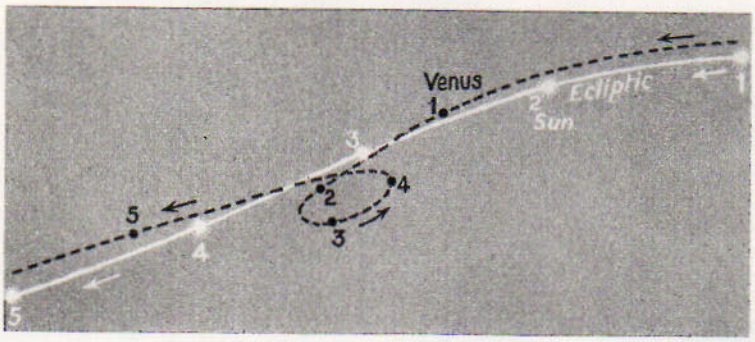


Fig. 146 — EXAMPLE OF A LOOP IN THE APPARENT PATH OF A PLANET (STATIONARY POINT AND RETROGRADATION)

of the planet — is shown in Fig. 145, as a path formed of loops, stationary points and retrogradations, as in the case above.

This is what we observe when a planet is at superior conjunction, for example. Unlike Mercury and Venus, it begins to appear in the east out of the morning twilight. Although retreating further from the Sun, it moves round the sky in the same direction as the latter. When it reaches quadrature, its apparent movement begins to slow down and then stops altogether, for the Earth is moving directly towards it and therefore appears stationary. Then as we begin to overtake it, our greater speed makes the planet seem to drift backwards against the stars. Finally, as all three bodies — Earth, Sun and planet — come into a straight line, we see the planet in opposition. At this time the difference between the speeds

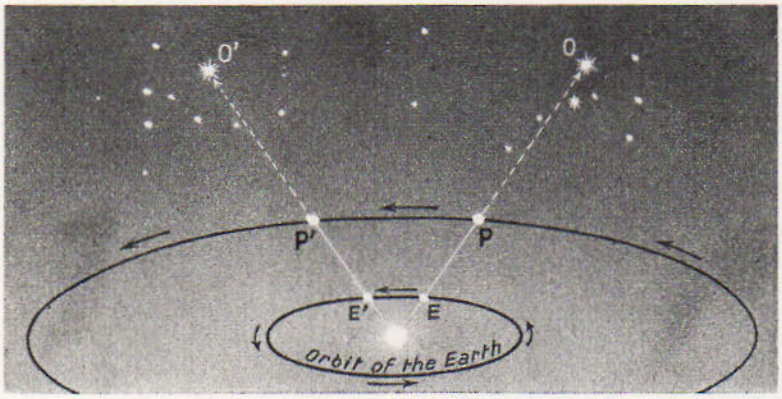


Fig. 147 — SUCCESSIVE OPPOSITIONS OF A SUPERIOR PLANET. When at P it will be seen against the stars at Q; when at P', it will be seen at O'.

of the two planets, which are moving on almost parallel courses, reaches its maximum value, and so does the apparent retrograde movement. But now, as we overtake and recede from it, the other planet seems to slow down, stop completely once more at eastern quadrature, then move on again straight towards superior conjunction. A new cycle then begins.

A planet's apparent movement against the constellations is determined, as we have seen, by its distance and speed, and the other elements of its orbit. Each of the stages defined above will be repeated only after a period equal to the Earth's revolution period, plus the time needed by the Earth to catch up with the planet, which will have moved on along its orbit, thus bringing them back to the same relative position. The time taken by a planet to return to the same relative position is called its synodic period. Thus the time between two successive oppositions increases the nearer the planet is to the Earth, because it is going faster and we take longer to catch it up.

The importance of this effect can be judged when we learn that oppositions of Mars occur, on average, every 787 days, whilst oppositions of Neptune are only a year and two days apart. From this it can be seen that, according to its apparent speed, a planet may be seen against one constellation at one opposition and against another constellation at the next. Thus it will gradually move round the Zodiac, going through all the variations in altitude that the Sun undergoes during the course of a year. Each planet has its own peculiarities of behaviour which affect the conditions under which it can be observed from the Earth. This applies particularly to Mercury and Venus, and will be discussed in the appropriate chapter.

Having now made our general survey of the Solar System and the behaviour of its members as seen from the Earth, let us go on to make a closer acquaintance with our neighbour worlds. Before tackling this exciting task, however, we ought first to learn more about our own world.

Although the Earth enjoys no special privilege as a member of the Solar System, we cannot help but consider it the most important: first because it is our home, and the conditions fostering organic life, the complex physical processes of the globe, are matters of supreme importance for us; and secondly, because it is the only heavenly body we can study first hand, that we can touch and experiment with. This is why the knowledge we gain about our native planet is not merely of importance for our daily needs but constitutes a vital element of comparison in the observations we make of other worlds than ours.

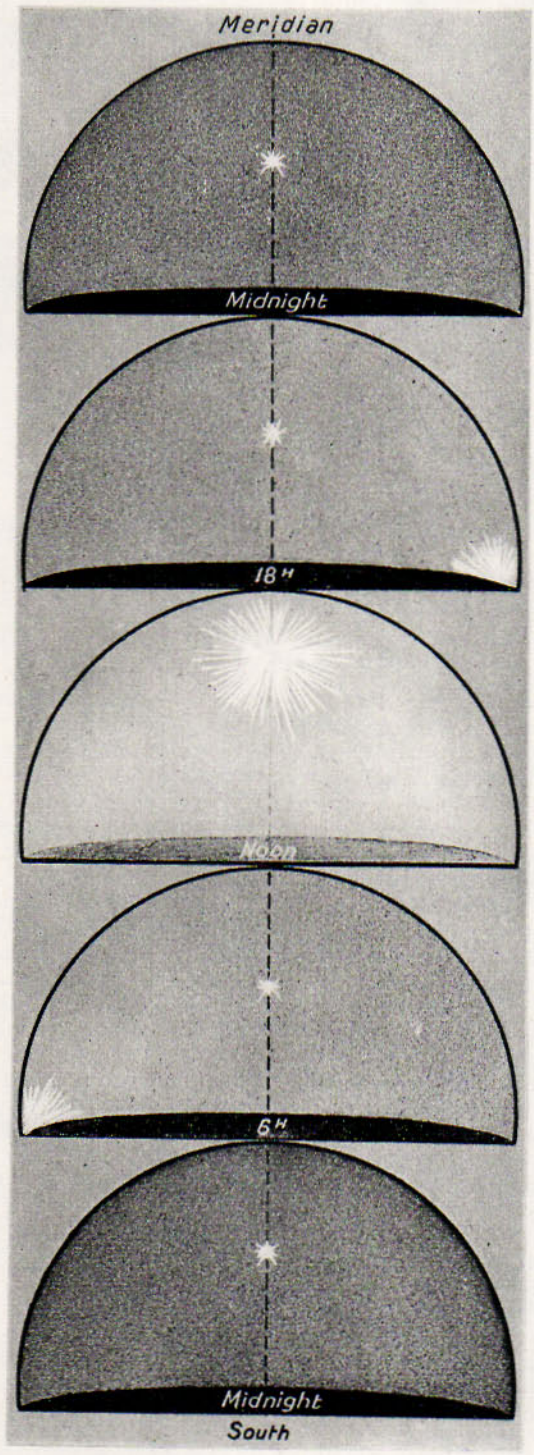


Fig. 148 — SUCCESSIVE POSITIONS OF A PLANET IN RELATION TO THE SUN

Fig. 149 — EARTH, SEA AND SKY

CHAPTER 4

THE EARTH AS A PLANET

Generally speaking we can consider the Earth as a sphere; at any rate, this is how it would appear to a distant observer on the Moon, for example. We shall discuss the regularity of the Earth's form later, but for the moment let us continue to think of it as a perfect sphere, as in preceding chapters. We have seen, too, how the different aspects of the sky and the alternations of night and day result from our position on the surface of the globe, the centre of which is the mathematical point to which all movements are referred.

We must now attempt to establish the relationship between two places on the surface of the globe as determined by the Earth's curvature. For this purpose we shall take its diameter as 7,927 miles. As will be seen, the size of the Earth is enormous compared to, say, a man. Furthermore, at the place where we are standing, the Earth's surface — the imaginary surface with all topographical irregularities smoothed out — gives the impression of being flat, at least over visible distance. For example, it is usual to take a stretch of water as representing a flat, horizontal plane. This impression would be supported by the use of a spirit-level floating on the water. But the horizontal determined in this way is simply a tangent to the globe, which, by reason of its large diameter, appears flat on the scale on which we see it. If we repeated this experiment somewhere else, we should again appear to have a flat horizontal surface, but if the two lines were projected towards each other, they would meet at an angle (Fig. 150). In fact a large expanse, such as a lake or sea, may be called 'flat', if by this we mean that all points on it are at equal distances from the Earth's centre.

This effect is shown in another way. If we fix the vertical of a place by the classical means of a lump of lead on the end of a piece of string, it will make an angle with a similar vertical set up elsewhere. The plumbline determines the vertical because of its weight; it is attracted by the Earth, which, as we have shown, behaves as if all its mass were concentrated at the centre, and this attraction brings the plumb-line into line with a terrestrial radius. Since the Earth is a sphere, the various 'vertical' directions found at different places all converge on the centre, like the spokes of a wheel.

Strictly speaking, therefore, even the smallest expanse of liquid is not flat, the most carefully levelled bowling-green is always curved, and the sides of a building, which the builder has taken such great care to make vertical with a plumb-line, are not parallel, but diverge from the ground upwards.

What will be the effect of the Earth's curvature when studied over distances large in comparison with its diameter? All around the Earth, the verticals of every place

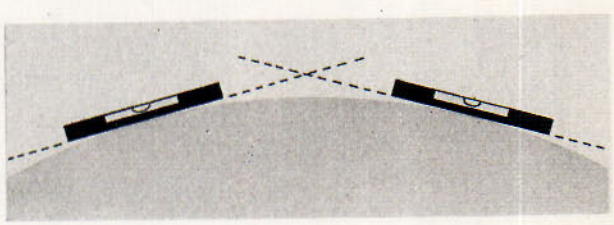


Fig. 150 — THE HORIZONTAL, AS SHOWN BY SPIRIT-LEVEL, VARIES FROM PLACE TO PLACE

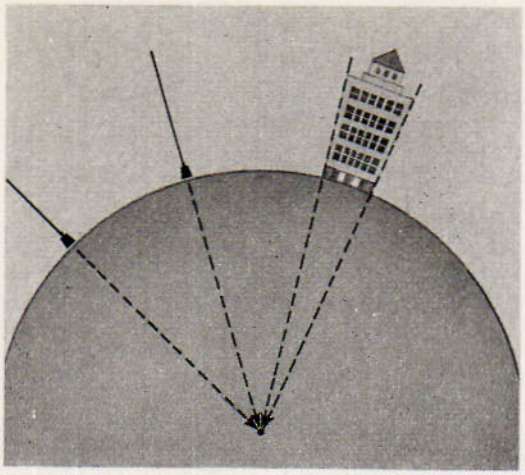


Fig. 151 — VERTICALS MEET AT THE CENTRE OF THE EARTH

on the surface of the globe radiate out into space, and this means that the projections of two ends of a diameter point in opposite directions on the celestial sphere. It is the same for us human beings, who stand on the Earth at all angles compared with one another, kept there by the force of gravity. From this we have developed the idea of 'up' and 'down', 'down' being the direction of the force of gravity towards the centre of the celestial (and terrestrial) sphere.

Consider two men standing on exactly opposite sides of the globe, so that each is at the antipodal point of the other. Their feet, although being 'down' for each individual, will in fact, be pointing towards the feet of the other person, whilst their heads will be pointing in opposite directions. We cannot experience this condition directly, but let us imagine our two men placed on a tiny world hardly bigger than themselves, but having a gravity sufficiently high to keep pulling them against the surface. Their relative positions would then be seen with ease

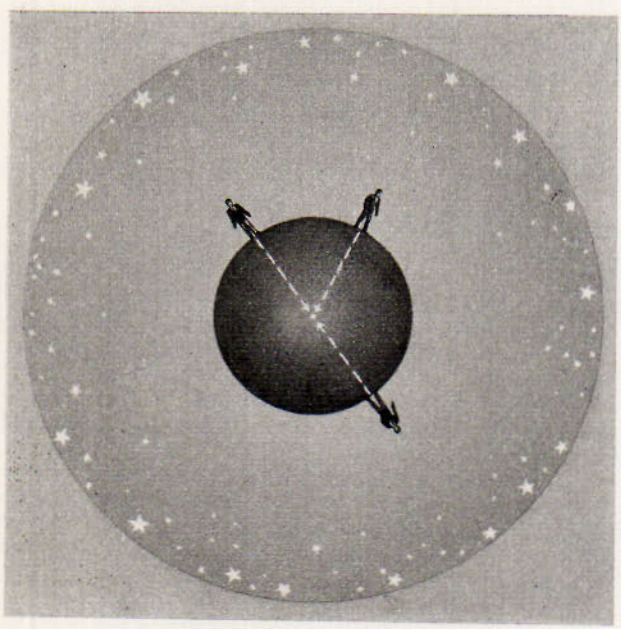


Fig. 152 — ORIENTATION OF A MAN IN DIFFERENT PARTS OF THE WORLD

(Fig. 153). This little demonstration shows that the notions of 'up' and 'down' are relative, valid only for the spot on the globe where we happen to be, and have no significance as far as the stars and space generally are concerned. No matter where we are, we all see the stars above us, but an individual star will be seen at an angle determined by our position on the globe: for one person it may be at the zenith, whilst for another, somewhere else, it may lie in exactly the opposite direction.

This will recall what we said in earlier chapters about the visibility of the sky, which depends on the situation of the observer and the rotation of the Earth. An observer would have to travel completely round the circumference of the Earth in order to appreciate the full effects of its curvature.

However, there are ways of demonstrating the curvature of the Earth more easily, and these are connected with the horizon.

The apparent and the astronomical horizon. — A provisional definition of the horizon was given in the section dealing with the appearance of the sky. We shall now discuss it in more detail, since it has a practical importance also.

The word 'horizon' is derived from the Greek *orizo*, bound or delimit. In fact we apply the term to the line all around us, where the land or sea meets the sky. Because of their apparent diurnal rotation, the stars move in relation to the horizon; their rising and setting actually mean that they are coming into view or disappearing round the edge of that part of the terrestrial globe which is visible to us.

In astronomical works the Earth is shown as a sphere, and the eye of the observer is assumed to be at the surface level of the globe. If this were so, the observer would be quite unable to see out over the surface, since, theoretically, the curvature of the globe means that the ground would fall away from him, immediately beneath a horizontal plane tangent at that spot. This plane is the astronomical horizon, and it is merely a boundary line; the half of the celestial sphere which happens to be above it is visible, while the other is invisible. This is valid whatever the observer's position on the globe, since the globe itself is absolutely insignificant in size compared with the celestial sphere which we may regard as infinite. The astronomical horizon marks the rising and the setting of the stars as calculated for any particular place, but in practice the observed times of sunset and sunrise will be slightly different from the theoretical prediction, since our observations are determined by the apparent horizon, and its position in relation to the heavens varies according to the situation of the observer.

Fig. 153 — EFFECTS OF PERSPECTIVE ON A SMALL GLOBE

Fig. 154 — THEORETICAL DEFINITION OF THE ASTRONOMICAL HORIZON

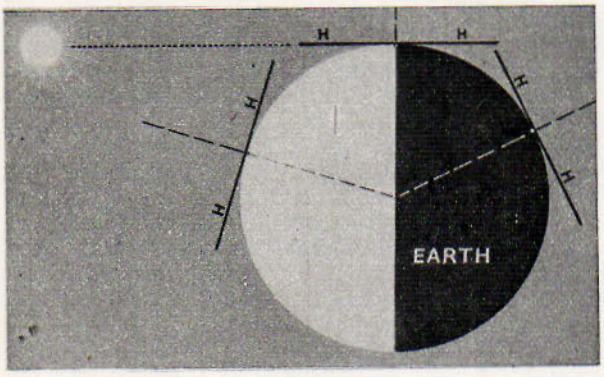


Fig. 155 — ASTRONOMICAL HORIZON (H) AT DIFFERENT PLACES ON THE TERRESTRIAL GLOBE

In fact, the observer is never placed where he is theoretically assumed to be in order to show the astronomical horizon, but is always at some height above the surface of the sea or land. Thus he overlooks the curved surface around him, and his horizon is where his line of sight just makes contact with the edge of the globe. This limit can be defined as the circumference of the base of a cone at the apex of which the observer's eye is placed, and is found to be beneath the astronomical horizon of the place.

This difference can be verified very simply by the careful use of a small telescope. The principle is as follows: a telescope can be set perfectly horizontal with the aid of a good spirit-level. When it is arranged in this way, it is obvious that the telescope does not point exactly at the apparent horizon, and in order to make it do so, the instrument must be pointed downwards by a certain amount (Fig. 157). The value of this angle is called the horizontal depression. It increases as the point of observation rises in altitude, as will be seen from Fig. 158. This is the reason for the above-mentioned difference between the sunrise as observed on the apparent horizon and as calculated for the astronomical horizon.

It is important to note, however, that in order to perform this experiment successfully, it must be made over a landscape which approximates as closely as possible to the geometrically flat surface shown in diagrams. Unfortunately the ground is rarely, if ever, as flat as this, and what we call the horizon is very often merely the line traced by innumerable irregularities differing in size and shape, and often seen in perspective one behind the other. Thus there may be quite important differences between the distances to the actual horizon, and the theoretical one assuming Earth to be a smooth sphere. In any case, the line joining earth and sky will be raised above the position it should occupy, and there will be a considerable variation in the times of the appearance and disappearance

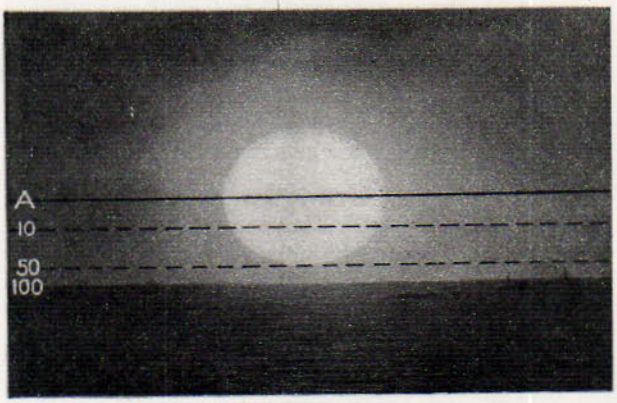


Fig. 158 — VALUE OF THE HORIZONTAL DEPRESSION COMPARED WITH THE SOLAR DISC, FOR DIFFERENT ALTITUDES. A, astronomical horizon.

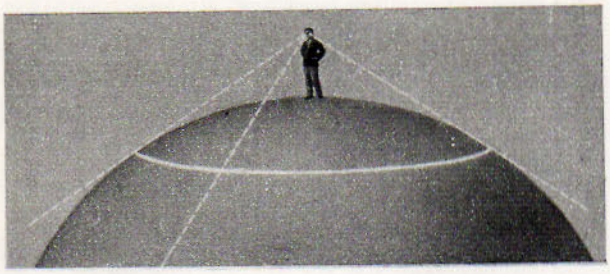


Fig. 156 — DEFINITION OF THE APPARENT HORIZON

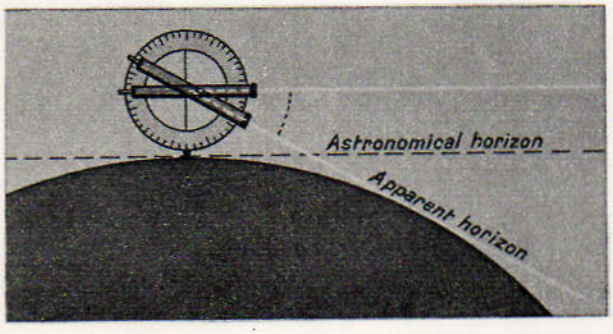


Fig. 157 — MEASUREMENT OF THE HORIZONTAL DEPRESSION

of the Sun compared with the calculated times for the astronomical horizon.

In contrast to this, the surface of the sea is a surface of equilibrium; that is to say, it follows the curvature of the Earth, and has therefore been chosen as a reference surface for the latter. The sea realises the theoretical conditions of uniformity, at least over large areas (for, strictly speaking, water is never in a state of absolute rest).

Waves and swells, however small, will always tend to mar the perfect geometrical smoothness of the sea, and over a short distance the importance of these small waves is liable to be preponderant. We may immerse our eyes in the water in order to be exactly at sea-level, but we cannot count on this to afford good conditions for making scientific observations; the slightest undulations in the water will take on an overwhelmingly false importance, and will prevent our making measurements. This may be observed if we look out at sea-level, when we shall see a very close horizon in no way corresponding to the one we ought to see; it is simply an irregular line, tossing and breaking with the waves which form it.

These details are not without a certain interest, and we would do well to dwell on them a little. When we wish to make observations and measurements, we must refer, in spite of everything, to a marine horizon, for only then

Fig. 159 — NO HORIZON. Houses and trees form an irregular screen in front of the observer.

Fig. 160 — BROKEN HORIZON. The skyline does not correspond with the theoretical horizon of the place.

Fig. 161 — SEA HORIZON. This is effectively the true geometrical horizon.

Fig. 162 — APPARENT HORIZON, standing out against higher ground beyond the limit of the theoretical horizon.

do the above-mentioned irregularities become negligible. Thus the edge of the ocean is the best place to observe the curvature of the Earth, for one can usually observe from both the shore and the top of the cliffs. Indeed there is no need to make exact measurements to detect this curvature: on a clear day, a sufficient number of well-spaced reference points such as islands and rocks, etc., will make it obvious.

In this connection we may recall the classic example of a boat approaching the coast from the open sea. The tips of the masts are seen first, on the horizon, then the topsails come into view, whilst the hull remains invisible. When the hull finally appears, it seems almost to be ascending a slope. The ship, when it is fully visible, seems poised on the horizon. Then, as it comes nearer, it seems to follow a curved path down from the horizon until it arrives in front of us.

Although this demonstration is theoretically simple, in practice it is more difficult than it seems. First of all, the ship must be in exactly the right position at the right time; the weather must be fine, and the air must be perfectly clear, so that the horizon is sharply defined. When we come to study refraction we shall see how different layers of air at the horizon can play the most curious and deceptive tricks with an image. Deformation or instability of images caused by waves in the air often render the smaller details unobservable, and this applies particularly in the present case.

Another more general source of information is connected with the height of the observer. If we survey a seascape with an indented coastline receding into the distance, or better still with a number of islands dotted here and there, we shall clearly see differences in how they appear, due to the curvature of the Earth. As we climb to successively higher vantage points, details which were hidden in the first place will begin to appear, and beyond them, in the far distance, we shall see the horizon. This

Fig. 163 — EXACTLY AT SEA LEVEL, THE HORIZON IS NOT SHARPLY DEFINED

may well prompt us to wonder how the distance to the horizon varies with the altitude of the observer.

For low altitudes, we can make a very simple calculation. Take the square root of the height of the eye above sea-level in feet, multiply it by 5 and divide the answer by 4. The result is the distance in miles to the apparent horizon. For example, we are standing on a cliff 100 feet high; the square root of 100 is 10; and if we multiply by 5 and divide the answer by 4, we get 12.5. The visible horizon is therefore 12.5 miles away.

When the observer's altitude is very high, this method becomes invalid, and it is necessary to make a more exact calculation for the angles shown in Fig. 165. An examination of the values for the horizontal distance as a function of the observer's altitude (*see* table) will show that the horizon recedes very rapidly at first as the altitude of the observer begins to increase; then the rate of recession becomes very much slower. We shall take two examples: if we rise from ground level to a height of 1,000 metres, the horizon rises from nil to 113 kilometres away, whilst for the same increase in altitude starting

Fig. 164 — A SLIGHT INCREASE IN HEIGHT ABOVE SEA LEVEL GIVES A MUCH WIDER HORIZON

from a great height, say between 9,000 and 10,000 metres, the distance to the horizon will increase only from 338 to 357 kilometres. If we continue to travel away from the Earth, as shown in Fig. 165, the change of angle of the line joining the eye to the horizon, which is a tangent, would be very slow, and we should reach a stage where we could double our distance without perceptibly increasing the amount of the Earth seen. At a very great distance — at about 30,000 kilometers and beyond — we should be able to see almost a complete diameter of the Earth, and therefore half of its surface, which is how we observe the planets in space.

The true shape of the Earth, therefore, is visible only from a very great height above the surface. In the above cases we have been making observations which confirm this by means of the observed effects. This has been necessary because we are accustomed to living fairly close to the surface of our planet. But now modern science has given men a more complete and exact view of their planet.

During balloon ascents, as pioneered by Professor Piccard, men have been able to attain heights of as much as 100,000 feet above the ground; at this height the curvature of the Earth begins to become apparent. Photographs from unmanned high-altitude rockets give an even clearer picture of the curvature (Fig. 167).

In this connection the rôle of the atmosphere becomes more than usually important. With such great horizon distances, the effects of refraction can become quite large, and they may even bring into view objects which lie below the geometrical horizon and which should therefore be invisible. This occurs because the light is forced to follow a slightly curved path round the surface of the globe. This means that the real horizon is further away than the geometrical horizon. The effects of refraction must be allowed for when calculating the horizontal depression.

The table, which allows for this effect, gives horizon distances (h) for various altitudes (a) up to a height of 5 kilometres, which completely contains the region of the atmosphere inhabited by human beings.

This account of the Earth's spherical nature could not be ended without some discussion of a question interesting from the tourist's point of view: how far away can we make out objects on the surface of the Earth?

Visibility at great distances. — In the above discussion of the distance to the horizon, the Earth was considered as a smooth sphere. When the line of sight just touches the sphere as a tangent, the distance at which it does so determines the limit at which the Earth's curvature cuts off our view. As we have seen, this is best shown over the sea, for the land horizon is usually irregular, and in mountainous country the term 'horizon' loses all meaning.

Although in many cases the topography of the location may restrict our view, it may also mean that we can see objects much further away than we should otherwise do

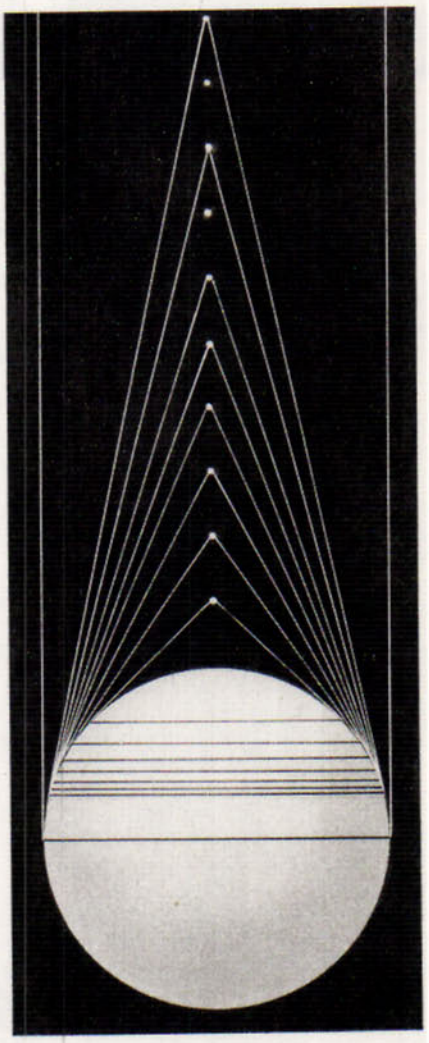


Fig. 165 — THE RATE AT WHICH THE HORIZON WIDENS DECREASES AS WE GET FURTHER AND FURTHER AWAY FROM THE EARTH

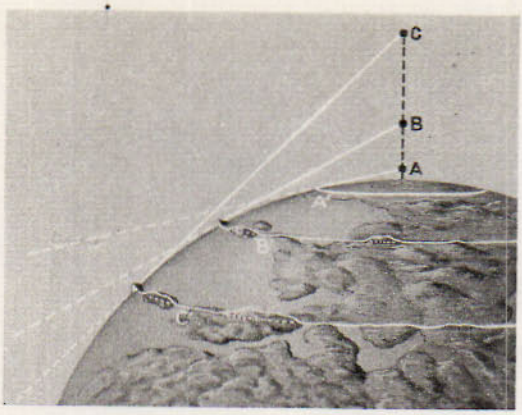


Fig. 166 — DISTANCE OF THE HORIZON AS A FUNCTION OF ALTITUDE

a = ALTITUDE, h = HORIZON DISTANCE

a	h	a	h	a	h	a	h
m	km	m	km	m	km	m	km
1	3.9	15	15.1	100	39.0	700	103
2	5.5	20	17.4	125	43.6	800	110
3	6.8	25	19.5	150	47.8	900	117
4	7.8	30	21.4	175	51.6	1000	123
5	8.7	40	24.6	200	55.1	1200	135
6	9.6	50	27.6	250	61.6	1500	151
7	10.3	60	30.2	300	67.6	2000	174
8	11.0	70	32.6	400	78	3000	214
9	11.7	80	34.9	500	87	4000	246
10	12.3	90	37.0	600	96	5000	276

on a smooth sphere. It may happen that the tops of mountains in the far distance show above the real horizon, whilst their lower slopes remain hidden by the curvature of the Earth. This is illustrated in Fig. 168. Without going into further details, we shall now consider a few interesting examples of this sort of occurrence.

From Monaco and Nice, it is possible to see the mountains of Corsica 120 miles away.

From Malta, one can see Etna, the famous volcano in Sicily, which is 7,591 feet high and 135 miles away.

From Bergamo (Italy) Mount Viso in the Alps can be seen at a distance of 145 miles.

From Notre-Dame-de-la-Garde at Marseilles, it is possible to see the Canigou (9,131 feet) jutting up from beneath the horizon, silhouetted against the setting sun; this occurs every year between 10 and 13 February, and 28 and 31 October. The distance is 160 miles. In this connection, we must remember what has been said about refraction, which may bend light from a point below the horizon round the surface of the Earth into our view, thus increasing our range of visibility. The visibility of the Canigou is an example, for, strictly speaking, it should not be visible from Marseilles: a straight line joining these two points passes, at its lowest, 390 feet beneath the surface of the Mediterranean.

Spain and Algeria have been linked for geophysical purposes by light signals travelling a distance of 175 miles over the sea between them.

Finally, from the summit of Mont Blanc (15,781 feet),

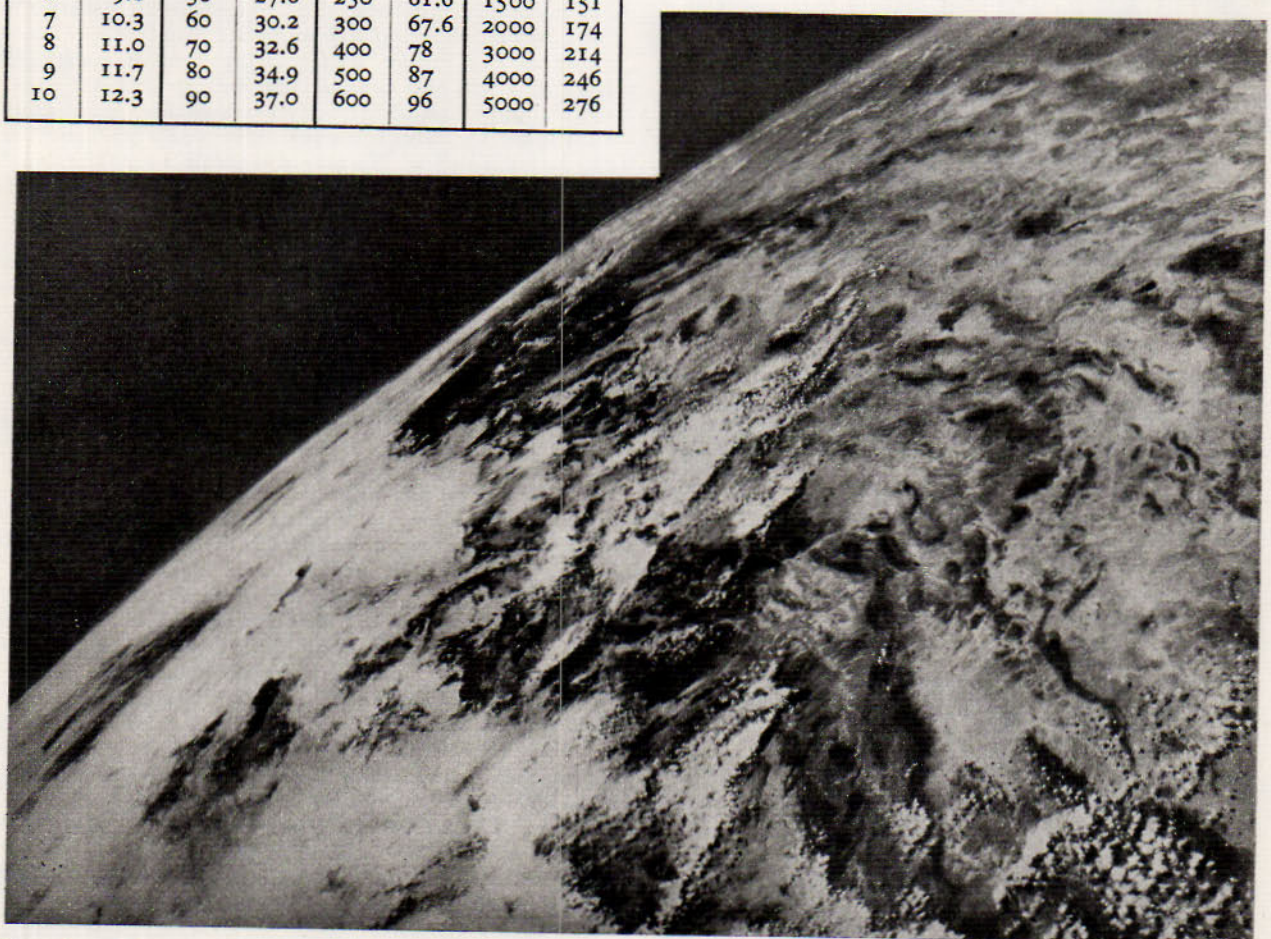


Fig. 167 — THE CURVATURE OF THE EARTH. A composite picture — the left-hand section from a height of 155 miles and the right-hand section at 138 miles — taken from a US Navy Viking II rocket. The horizon is about 1,100 miles from the camera. *(Keystone photo)*

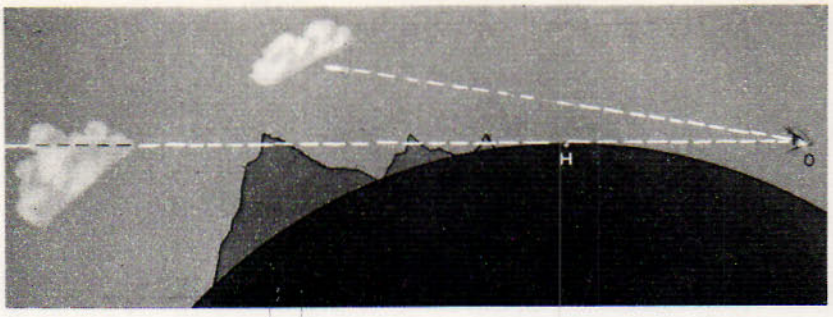


Fig. 168 — FOR AN OBSERVER AT O, THE HORIZON (H) MAY BE FORMED OF OBJECTS AT VERY DIFFERENT DISTANCES

it is possible to see places as far as 210 miles away in the right conditions.

Observations such as these require very pure atmospheric conditions and the right illumination, as in the case of the Canigou silhouetted against the setting sun.

Like mountain tops, high clouds may also be seen at a very great distance; this may be very useful to the meteorologist, as it enables him to obtain information about the state of the atmosphere over a very wide area.

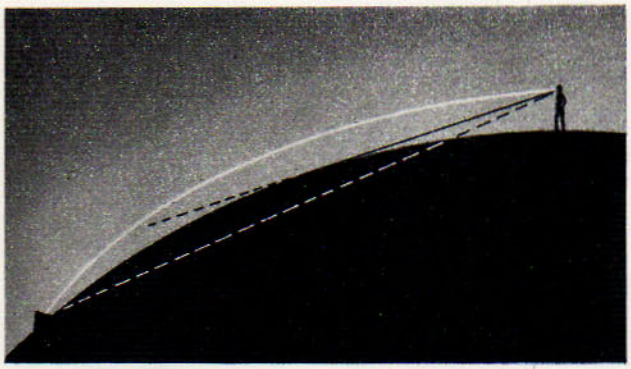


Fig. 169 — ATMOSPHERIC REFRACTION BRINGS INTO SIGHT OBJECTS BELOW THE HORIZON

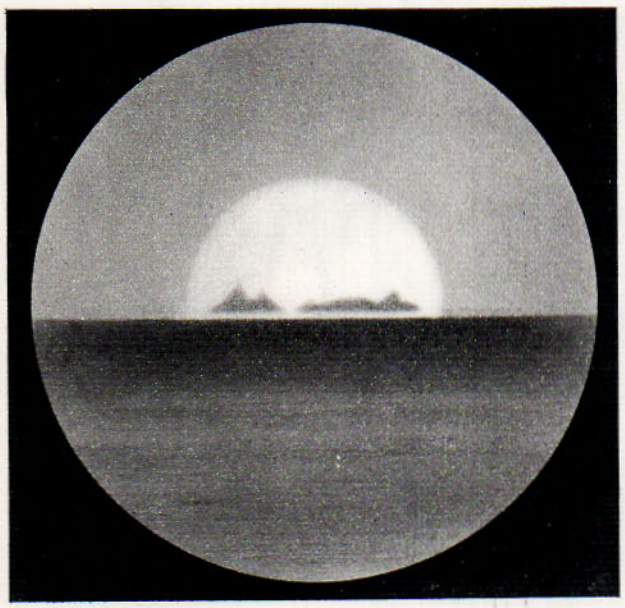


Fig. 170 — THE CANIGOU, SEEN FROM MARSEILLES, SILHOUETTED AGAINST THE SETTING SUN, 12 February 1905.

The exact shape of the Earth. — In earlier chapters we saw how fairly simple methods of observation not only enabled us to detect the curvature of the Earth, but also gave us a good estimate of its size. However, approximations such as these are not good enough for most astronomical or practical purposes. Later we shall learn how much more refined our ways of measuring the Earth have developed through the years, but, for the moment, we are shall consider why it can no longer be regarded as a perfect sphere.

The Earth is a sphere flattened at the poles and with a bulge at the equator; this figure is called an ellipsoid of revolution.

In order to account for its present shape, we must call upon what is known as 'centrifugal force', and we must also make the hypothesis that in its earlier history the Earth was much nearer the fluid state than it is now. Any object on the surface of a rotating body is constantly liable to fly off at a tangent, and will do so if the gravitational attraction of the body is insufficient to maintain it at the surface. This is the principle used when throwing a stone from a sling. We may also recall those fairground shows which apply the principle with a success which may be gauged by the enthusiasm of the public. People stand on a platform which begins to spin round faster and faster; there comes a time when the friction between them and the floor of the platform (due to their weight) is no longer sufficient to resist the centrifugal force, which increases with the speed of rotation. For a given rotational movement, when two points are at unequal distances from the centre of rotation, the one which is further away, having to cover the greater angular distance in the same time, will have the greater tendency to fly off at a tangent.

According to this principle, it is easy to study how a rotating sphere of some liquid material would behave if every part of it were free to obey the forces acting upon it. If the sphere were perfect when motionless, it could not remain so when rotating. Let us consider a point A on its surface (Fig. 172). At the same time as it is attracted along AO by the action of gravity AG towards the centre of the mass, the centrifugal force which arises from the speed, proportional to the circumference which it follows, tends to expel it perpendicularly to the axis of rotation, along AF. This point will therefore be subjected to a force AG′ (the resultant of the two forces acting simultaneously) which does not pass through the centre of the globe. Centrifugal force increases from the pole to the equator, where the speed of the surface is at its maximum. This increasing force acts on the molecules of the fluid globe in question. If they can freely obey all the forces acting upon them, they will take up a shape which is called a figure of equilibrium, everywhere obeying the resultant force AG′. Mathematical analysis shows that this surface must be an ellipsoid of revolution, that is to

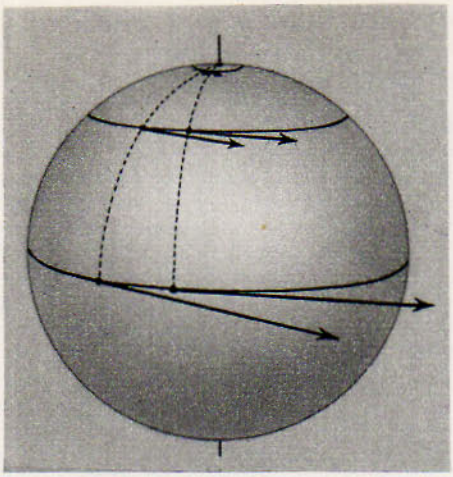


Fig. 171 — THE EARTH'S SPIN TRIES TO THROW OFF OBJECTS AT THE SURFACE. This force — centrifugal force — is strongest at the equator and weakest at the poles; this gives rise to the Earth's equatorial bulge and polar flattening.

say, the solid traced out by an ellipse turning about its minor axis.

To put the matter more simply: we start with a given mass (which will remain constant), in spherical form, and cause it to rotate. The equatorial region will bulge out under the influence of centrifugal force, to the detriment of the polar regions, which will be flattened in proportion.

It must be said that the oblateness of the Earth is not so marked as would appear from the accompanying diagram (Fig. 172), which is exaggerated for clarity. In fact, the flattening of our globe can be detected only by very refined methods of measuring, which is why we said that a distant observer would see the Earth as an apparently perfect sphere, and why it is permissible to simplify in this way for general purposes. It would be quite different if our planet were otherwise constituted or rotated faster. Both Jupiter and Saturn turn much more rapidly than our world (in approximately 10 hours instead of 24 hours), and moreover, they are composed of much lighter materials. Even a small telescope will show the flattening of these two planets, which in the case of Jupiter amounts to 1/15 and of Saturn 1/10, whilst the

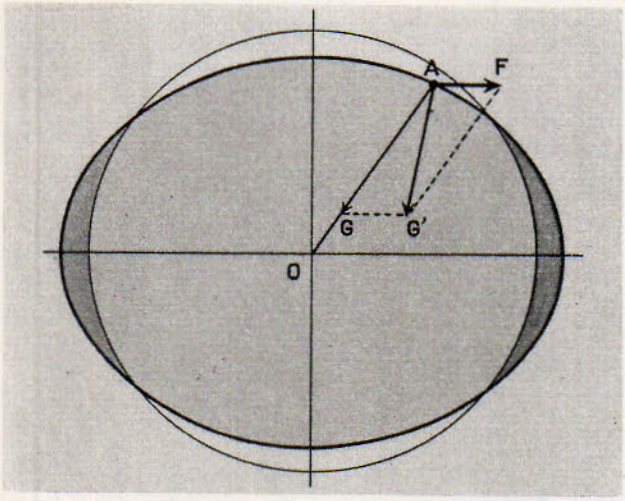


Fig. 172 — COMPOSITION OF THE FORCES WHICH HAVE DETERMINED THE FLATTENING OF THE TERRESTRIAL GLOBE

flattening of the Earth is only 1/297. These figures represent the relation between the major axis and the difference between the two axes. Taking the major axis as a and the minor axis as b, the amount of flattening is given by $\frac{a}{a-b}$.

The size of the Earth. — The methods used to discover the flattening of the globe needed much painstaking work to yield results. Different values have been obtained at different times, but the most accurate appear to be those arrived at by Hayford in 1909, and since adopted internationally for use in all data concerning the form of the Earth.

Oblateness	1/297
Equatorial Radius (a)	6,378,388 metres
Polar Radius (b)	6,356,912 metres
$a-b$, or oblateness, in metres	21,476 metres
Length of elliptical meridian passing through poles	40,009,152 metres
Length of Equator	40,076,594 metres
Surface Area	510,101,000 sq. kms.
Volume	1,083,320,000,000 cu. kms.
Volume of sphere with radius equal to minor axis of ellipsoid	1,076,100,000,000 cu. kms.
Volume of equatorial bulge (difference between last two figs).	7,220,000,000 cu. kms.

If we relate one of these data to the spherical form, we get:

Radius of sphere with same meridian as Earth	6,367,654 metres
Radius of sphere with same area as Earth	6,371,228 metres
Radius of sphere with same volume as Earth	6,371,221 metres

Finally, one more point must be noted. Although the above data refer to an ellipsoid of revolution, this may not be the actual shape of the Earth. According to Helmert, we could postulate an ellipsoid with three unequal axes, two situated equatorially and nearer to each other than to the polar axis. It will be the task of future geophysicists to decide whether this is so.

Definition of the metre. — In many places the metre has been substituted for the older measures, its main advantage lying in the fact that it is a definite fraction of the Earth's diameter. The law promulgating the establishment of the metric system in France, passed on 7 April 1795, defines the metre as one ten-millionth part of a terrestrial quadrant, and the standard metre was established according to the values for the Earth's diameter current at that time. Since then, better geophysical methods have come into use, giving a more accurate estimate of the Earth's size. We must therefore confess that a ten-millionth part of a quarter-meridian is 0.00022 metre longer than the standard metre bar lodged in the National Archives of France.

The question was then whether to construct a new standard metre in the light of later knowledge or let the other one stand. As these methods are always subject to improvement, it seemed preferable to retain the earlier one, with the full knowledge of its exact error. The metre has therefore been re-defined: it is the length, at 0°C, of the platinum bar which has been adopted as the standard unit of length by international agreement, and which reposes in the International Bureau of Weights and Measures at Sévres, France. We, in this country, have been in

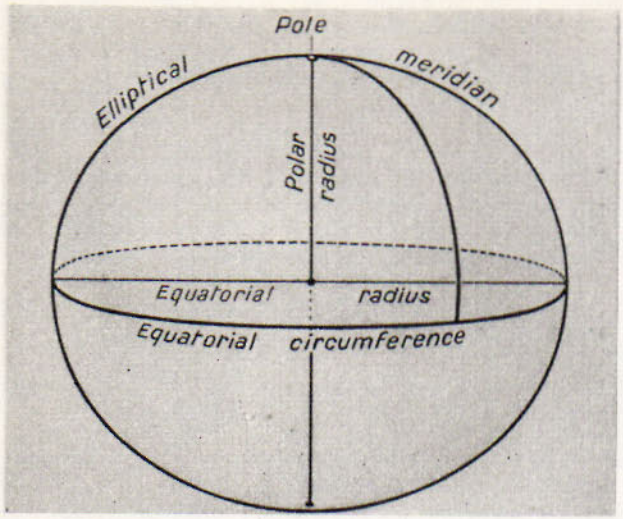


Fig. 173 — THE EARTH IS AN ELLIPSOID

the habit of using English measures in our daily lives, but the metric system is almost universally used for scientific purposes. In this book, therefore, the plan has been to employ the metric notation wherever possible, but where it is felt that English measures would be more easily understood, we have used them.

Lengths of meridian arcs and latitude circles. — For convenience in locating places on the surface of the Earth, the globe is divided by a series of imaginary circles of longitude and latitude, at right angles to each other, which form a system of co-ordinates.

The system of meridians, or great circles of longitude, begins at a conventional starting point. The parallel latitude circles are drawn between the equator and the poles. They are all divided into degrees, minutes and seconds of arc.

On a perfect sphere, an arc of 1° would embrace the same distance on the surface at any part of a meridian. From what we have said about the shape of the globe, it will be recalled that a meridian is an ellipse, not a circle. Consequently an arc of 1° in this ellipse has different values at different latitudes. Fig. 174 shows why this is so. Let us consider the point A on the surface of an ellipsoid,

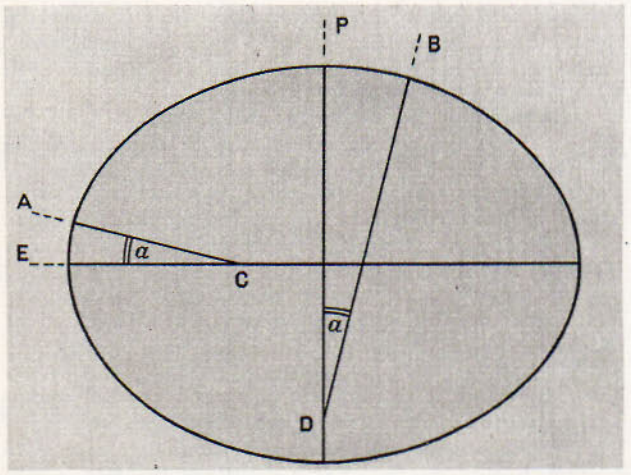


Fig. 174 — INCREASE IN THE LENGTH OF A DEGREE OF LATITUDE FROM THE EQUATOR TO THE POLE

whose equator is shown as E. The line perpendicular to the surface at this point meets E at C, making an angle a between them. In practice there is very little difference between the arc AE and a circle with centre C and radius CE. Let us take a similar situation with point B near the pole, P. Here, on account of the flattening, the perpendiculars will meet much further beneath the surface than in the first case, and for them to meet at the same angle, a, the distance between P and B on the surface must be larger than between A and E. Thus the same degree of arc will embrace a much larger distance on the surface than in the first case. The accompanying table gives the increasing values for the same meridional arc with increasing latitude.

Because of this variation in the value at different places, the nautical mile, the universal navigational standard, is a mean value. The first definition of a nautical mile corresponded to one minute of arc; this would range from 1,842 metres at the equator to 1,861 metres at the poles. The hydrographic conference of 1929 adopted a length of one minute of arc at latitude 45°, i. e. 1,852 metres.

The diameter of latitude circles decreases as we approach the pole. The length of a one degree arc at any

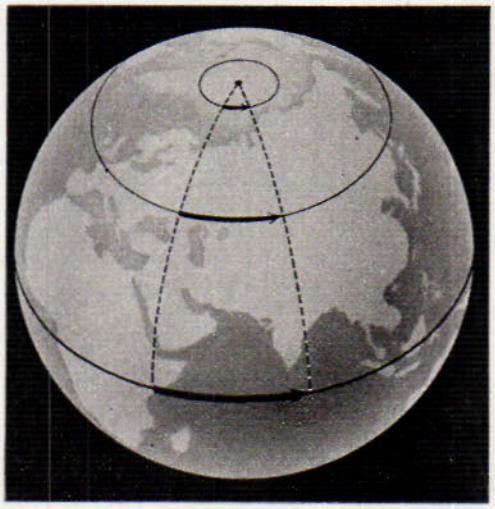


Fig. 175 — ROTATION OF THE EARTH. Distance travelled by different points on the surface during equal times.

parallel is a function of latitude; it is greatest at the equator and zero at the poles, which are only mathematical points. The table gives the corresponding lengths of the same arc on different parallels, at intervals of 10°. The time taken for the surface to move (by rotation) through this angle is given at the head of each column.

The speed of rotation at any point on the Earth's surface is determined by the latitude. This speed is zero at the poles, which may be thought of as points turning on themselves, and increases to a maximum at the equator, as shown in the table. As some indication, the following are the surface speeds at different latitudes:

Pole	0 metres per sec.
60°	233 metres per sec.
30°	403 metres per sec.
0°	465 metres per sec.

Even though we in Britain are travelling very rapidly, we do not have any sensation of movement. Whilst these speeds are quite high, as speeds go, they are not tremendously so, especially when compared with our orbital

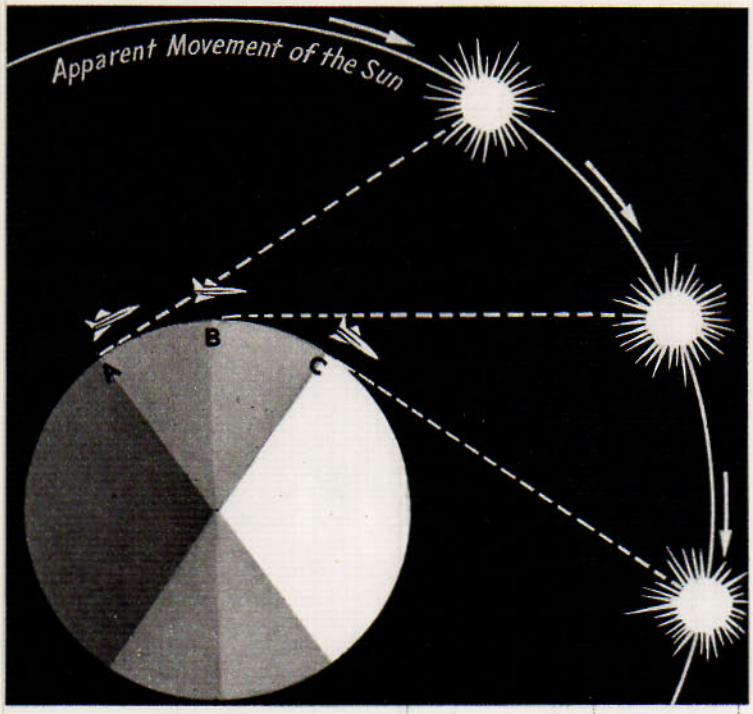


Fig. 176 — FOR AN AIRCRAFT FLYING AT THE SAME SPEED AS THE EARTH'S ROTATIONAL SPEED, BUT IN THE OPPOSITE DIRECTION, THE SUN WOULD APPEAR TO STAND STILL IN THE SKY

velocity round the Sun, for example. Indeed, man-made vehicles such as aircraft and rockets have already surpassed them. This means that if an aeroplane flies continuously into the sunset, at the right speed and keeping to the same parallel of latitude, the Sun will stand still on the horizon, because the plane's speed is equal and opposite to the speed of the ground by rotation; the Earth will then appear to drift past underneath the seemingly stationary plane (Fig. 176). At latitude 60°, the ground rotation speed is 525 miles an hour, which is well within the capacity of modern aircraft.

The mass of the Earth: force of gravity. — In the first chapter we warned against confusing the mass of a body with its weight, since the weight is the result of the mass; it must always be remembered that when an object is said to have a certain weight and falls to the ground when left unsupported, this is due to the attraction it is undergoing from the whole mass of the globe.

Strictly speaking, therefore, the Earth cannot have weight, only mass. With a volume of 1,083,320,000,000 cu. km. at an average density of 5.52, the Earth's mass is 5,977,000,000,000,000,000,000 metric tons, or 6×10^{27} grams.

In the following discussion, it should be remembered that the gravitational attraction of a certain mass upon a body at its surface acts as if the whole mass were concentrated at the centre.

It is not necessary to dwell on the subject of gravitational attraction and the acceleration of falling bodies etc., since such matters are adequately dealt with in the classical text-books of physics. Suffice to say that a body left unsupported will fall to the ground in a special way: at first its speed will be zero but then it will start to fall faster and faster. After one second, its speed will be numerically equal to the acceleration of gravity (g); after two seconds, the velocity is twice this (2g), and so on. A body will fall 4.9 metres (16 feet) in the first second,

ø	MERIDIANS OF LONGITUDE			PARALLELS OD LATITUDE			
	Arc of 1°	Arc of 1′	Arc of 1″	Arc of 1° described in 4m	Arc of 1′ described in 4s	Arc of 1″ described in 0.07s	Arc described in 0.1s
	metres	metres	metres	metres	metres	metres	metres
0°	110,576	1,842.90	30.71	111,324	1,855.40	30.91	46.38
10	110,609	1,843.48	30.73	109,644	1,827.40	30.46	45.68
20	110,706	1,845.10	30.75	104,651	1,744.18	29.07	43.61
30	110,855	1,847.57	30.79	96,490	1,608.16	26.80	40.20
40	111,038	1,850.63	30.84	85,398	1,423.30	23.72	35.58
50	111,233	1,853.88	30.90	71,699	1,194.98	19.92	29.81
60	111,417	1,856.94	30.95	55,803	930.05	15.50	23.25
70	111,567	1,859.43	30.99	38,188	636.46	10.61	15.91
80	111,666	1,861.09	31.09	19,395	323.25	5.39	8.08
90	111,700	1,861.66	31.03	0	0	0	0

19.63 metres (64 feet) after two seconds, and so on. An object dropped from the top of the Empire State Building (height 1,472 feet) would fall to the ground in about 9.36 seconds (Fig. 177).

If the Earth were stationary in space, the intensity of gravity would depend solely on the mass and radius of the globe, since the force decreases as the inverse square of the distance. But since the Earth is rotating, we must also consider the centrifugal force which this rotation generates, as we saw in our discussion of the Earth's flattening, and which acts against the direction of the central attraction. Thus the effective gravity is the resultant derived from the composition of these two opposing forces, and it varies from place to place according to the figure of the Earth. It may be deduced experimentally from the number of oscillations of a pendulum in a given time, there being a formula which gives the acceleration of gravity (g) from the length of the pendulum and the number of oscillations. The following values are derived in this way.

Lat.	g.
0° (Equator)	978.049 cm/sec²
51° 32′ (London)	981.183 cm/sec²
90° (Poles)	983.221 cm/sec²

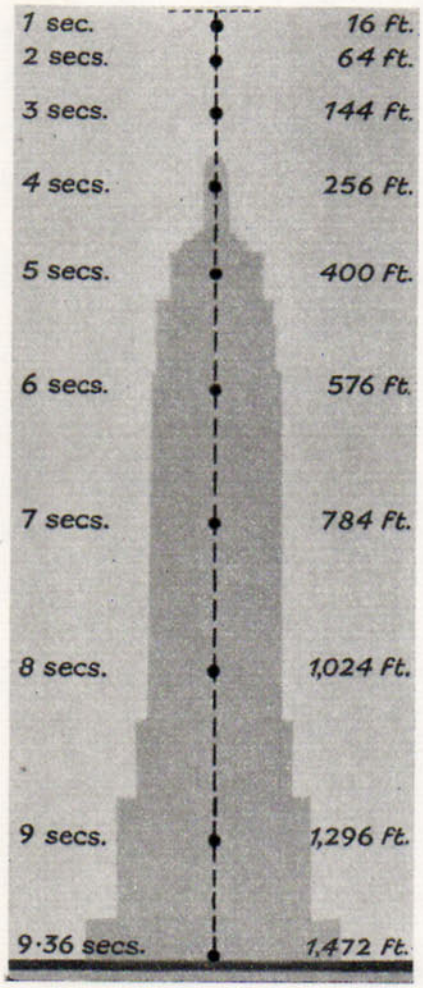


Fig. 177 — THE ACCELERATION OF GRAVITY AT THE SURFACE OF THE EARTH

For purposes of comparison these values have been reduced (that is to say, corrected) so as to relate them to the same surface level, the mean level of the oceans. In its prolongation beneath the surface of the continents, this level determines what we call the geoid.

The surface of the geoid corresponds to the constantly normal conditions established by the gravitational force-field for the whole of the non-homogenous globe. This special surface does not coincide exactly with that of the theoretical terrestrial ellipsoid, but is nevertheless a very close approximation.

It follows therefore that the weight of a body is not an invariable quantity, but depends on the locality where it is measured, i.e. the distance from the centre of the Earth. This cannot, of course, be demonstrated with a pair of scales, because the counterweight varies in the same way as the object being measured; instead, a very delicate spring balance called a dynamometer must be used. With this instrument an object weighing, say, exactly one kilogram at the foot of the Eiffel Tower would be found to weigh one gram less at the summit (300 metres higher).

When we go down a mine, we ought, theoretically, to detect a corresponding reduction in weight, since it is no longer the attraction of the whole mass we are considering but that of a sphere whose radius equals the distance between the bottom of the mine and the centre of the Earth. This however would hold only for a homogeneous sphere, which is not the case with the Earth. Repeated experiments in deep mines have, on the contrary, revealed a slight increase in weight, which means that the density of the deeper regions must be grater than that of the surface layers.

As will be seen later, the central regions probably consist of a core of very high density.

General Structure of the Earth: The oceans and continents. — After the rather dryly geometrical discussion of the last few sections about the Earth's size, shape and mass, we now go on to examine what might be called the physiognomy of our globe: the conditions and all the physical processes occurring on its surface.

In many ways an observer some way out in space would be better placed than ourselves, for he would have an overall view of certain large-scale features which we do not properly appreciate simply because we are too near to them.

We may divide the Earth's exterior into three basic parts: the atmosphere, the oceans and the continents.

We are familiar with the general appearance of the globe from the study of maps in our school-day geography lessons. What strikes us immediately is the preponderance of oceans over land areas. In fact, the area of the Earth covered by water amounts to 361,200,000 sq. km. whilst the dry-land area is only 148,900,000 sq. km.k

This relationship calls for comment. The diversity of coastlines is well-known; in many places they are smooth and flat, so that a change in level of a few dozen feet would inundate or uncover immense expanses of territory; on the other hand, where the coastline is steep and the land rises sharply to an appreciable height above the sea, an equivalent change in level would not produce anything like the same results; the water would rise as it does up the sides of a deep basin. This factor is very important in the behaviour of tides. Taking the world as a whole, however, any change in the sea level would lead to a considerable alteration in the present appearance of the land areas.

In any case, the enormous disproportion of the land and sea masses is very striking. Compared with the other planets, the Earth, as far as we know, is unique in this

respect. Where we can detect the presence of water on other worlds, it is in vastly smaller amounts than on the Earth.

Not only are the areas of the oceans and continents very unequal, but so is their distribution between the two hemispheres. In the northern hemisphere, the continental area only amounts to two-thirds of the sea area; in the southern, the difference is even more striking, as there is only one-fifth as much land as water.

Here we are thinking of the geographical hemispheres, but we can also divide the Earth so that one hemisphere contains the maximum land area, the other the maximum sea area. The late Professor Berget calculated that this would mean putting the 'land-pole' just off the coast of Brittany (*see* Fig. 179). In the land-hemisphere, land and sea are almost equal; in the other, the land is almost lost in the tremendous expanse of ocean.

This fact leads to the following conclusion: in nineteen cases out of twenty, a point situated on a land mass will have water at its antipodes. In more general fashion, any point on the Earth's crust projecting above the level of the sea is matched by a depression on the opposite side of the world. Thus we arrive at a sort of law establishing a diametrical opposition of the continents — or outcrops in the crust — and the oceans, which are depressions filled with water. A map of the world will quickly confirm this.

Most of the land masses of the Earth lie in the northern hemisphere, and they may be divided into three massifs: the Afro-European continent, (neglecting the Mediterranean depression), the Asian continent, with its south-western prolongation into Australia, and the American continent. Separated by the vast expanses of the Atlantic, Pacific and Indian Oceans, they form 3 triangular 'bosses' pointing towards the south pole. The latter is occupied by a vast continent, only now being explored,which contains some mountain ranges of considerable height. This constitutes a further 'boss' corresponding to the depression occupied by the frozen ocean round the north pole, where the ocean bed is more than 10,000 feet deep.

This arrangement of land and sea corresponds to a geometrical figure, the tetrahedron. The tetrahedron is a regular polygon with 4 sides in the shape of equilateral triangles; it thus comprises 4 corners, each opposite a side. Apart from the fact that the sides are not straight, the Earth, with its 4 bosses almost symmetrically distributed with respect to the ocean depressions, shows

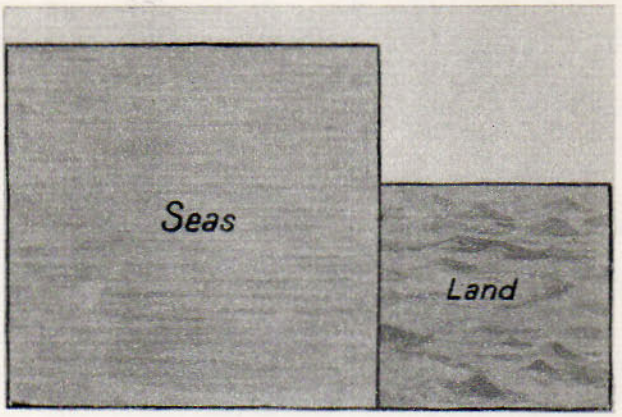


Fig. 178 — THE RELATIVE AREAS OF LAD AND SEA ON THE EARTH

a certain tendency towards tetrahedral form, its south pole being one of the points. We may explain this by assuming that in its early history, the Earth's crust contracted upon cooling.

It may be objected that this distorts the regularity of the ellipsoid mentioned in our earlier discussion. Theoretically, this is true, and it is an added complication to the work of the geophysicist. Practically speaking, these differences would remain quite unnoticed by our distant observer in space; indeed, he would not even be able to detect the flattening of the globe without very accurate measuring instruments, although the difference between the equatorial and polar diamteres amounts to 27 miles. When we remember that the highest points on the globe do not exceed 5 or 6 miles and that the mean height of the land masses is between 1 and 2 miles at the most, it will be seen that these little irregularities are truly minute in comparison with the diameter of the Earth, so that from a moderate distance out in space our planet would, at first glance, seem to be a perfect sphere.

In fact Mount Everest's height of 29,141 feet represents only one fourteen-hundredth part of the Earth's diameter. If we draw the globe 1 m. across, Everest, on the same scale, would be represented by a little wrinkle 0.7 mm. high. Thus it is wrong to compare our world with an orange, for the irregularities on the peel of an

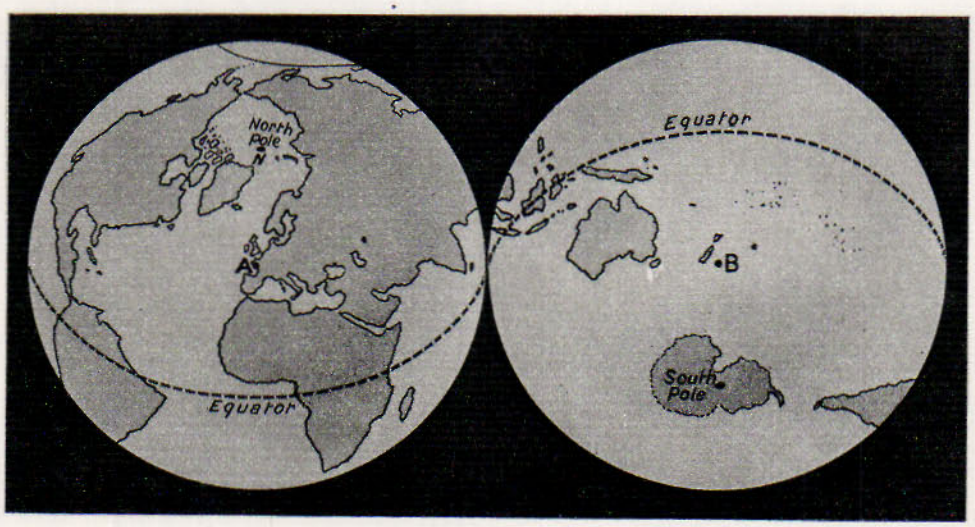


Fig. 179 — A, LAND POLE; B, OCEANIC POLE

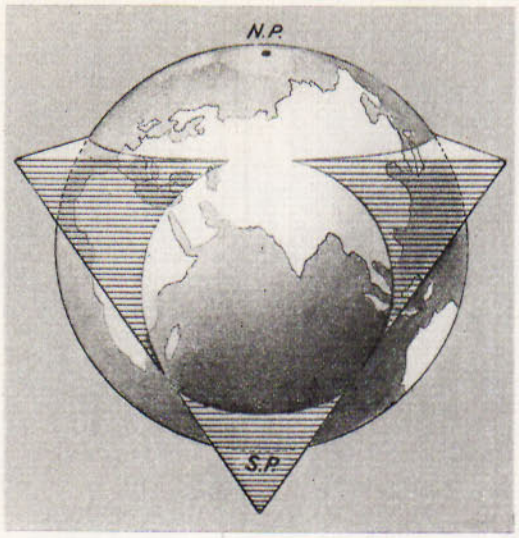


Fig. 180 — THE EARTH AS A TETRAHEDRON

orange are much larger, comparatively speaking, than the mountains on the Earth, which would have to be anything up to 280 miles high on the same scale.

It may be objected that we have considered only the continental irregularities on the Earth's surface, neglecting the ocean deeps, since, if we took these as our mean level, the mountains would be much higher. If we do this, then the difference between the top of Everest and the floor of the Mindanao Deep is 63,960 feet. In spite of this, an orange is still more rugged than the Earth.

The volume of the continents and oceans. — We have seen that compared with the diameter of the globe the height of the surface relief is insignificant. It must, however, be taken into account when studying the physical constitution of our world and when investigating its possible origin and course of evolution.

Let us therefore consider the total volume of water filling the ocean basins and the volume of land raised above the surface of the sea. We cannot state these volumes with the same confidence that we can state the corresponding areas, for it is difficult to allow for irregularities both on the land and beneath the sea, the ocean beds of the world not yet being completely explored.

In spite of this, however, the figures obtained are sufficiently impressive. The volume of the land masses above sea-level is estimated at 100,000,000 cu. km. and the volume of the seas at 1,300,000,000 cu. km. The disproportion is therefore greater than between the respective surface areas. The relation between the areas is 2.54 to 1, whereas between the volumes it is 13 to 1.

There is a way of illustrating this difference: if we levelled off all the irregularities on the surface of the Earth and spread out the material in a uniform layer across the whole land-surface of the globe, it would form a layer 2,275 feet thick. A similar operation to level the sea bottoms, keeping the areas of the basins the same, would make the average depth of the sea about 11,375 feet (see Fig. 185).

The atmosphere. — The atmosphere, the envelope of gas around the globe, is one of the most essential characteristics of our world, and for this reason we must consider it in the light of the effects it produces, as much as in and for itself.

We have already indicated its rôle in the diffusion of the Sun's light and the distortion and refraction of astronomical images. Its extension, composition and properties, not to mention the frequent disturbances which occur in it, are factors affecting our everyday life. The atmosphere, moreover, is the transition zone between the solid surface of the globe and the vacuum of space, and many of the phenomena occurring there are astronomical in origin.

For this reason meteorology, the study of the atmosphere with all its vicissitudes, largely depending on the behaviour of the Sun, is inseparable from astronomy, if only because the atmosphere constitutes a serious obstacle in the way of astronomical progress.

As we all know, the surrounding atmosphere forms a medium which is indispensable to he existence of organic life — human, animal and vegetable — and, indeed, the atmosphere of any planet is of prime importance. Conditions on the surface of any world are largely determined by the presence or absence of an atmosphere, and

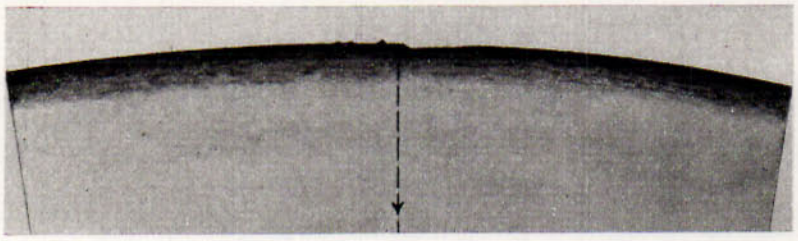


Fig. 181 — THE SIZE OF MOUNTAINS ON THE EARTH COMPARED WITH THE CURVATURE OF THE GLOBE

Fig. 182 — A SECTION THROUGH AN ORANGE, SHOWING THE SIZE OF THE IRREGULARITIES ON ITS PEEL

Fig. 183 — THE HIGHEST POINT ON EARTH: MOUNT EVEREST. *(Photo, Gaumont)*

its composition, if present; the deeper our knowledge of what goes on in our own atmosphere, the more accurately shall we be able to interpret the information coming throught it from the outside universe.

Thus the study of the atmosphere can be approached in two ways: we can investigate both its practical importance for life on the Earth and the contribution it can make to the understanding of many astronomical problems. Here, however, we do not propose to give a complete account of the laws governing meteorological phenomena, which can be studied to better advantage in a more specialised text; instead we shall confine ourselves to information necessary for the intimate understanding of our own planet, its origin and history.

Thickness of the atmosphere. — What we often simply call 'air' is in fact a mixture of many gases in different proportions. It envelops the Earth in a spherical mantle and although, as yet, we have not determined its full extent, it must extend several hundred miles above the surface. In this connection, the Earth's rotation imposes a certain theoretical limit beyond which there will be no remaining air; at this distance molecules of air would be spilled off into space at a tangent, overcoming the Earth's gravity. This limit is calculated to be at 24,560 miles above the surface. For the same reasons this sphere of air must be flattened even more than the globe itself, since its outermost parts, at whatever distance they may actually be, are further from the axis of rotation.

However, these are purely theoretical considerations, and the actual height of the atmosphere must certainly be less. Up to the present time, the question has not been settled.

It has long been known that the air becomes thinner the higher we climb; mountaineers and high-altitude flyers have neglected this to their cost. It is also known that at great heights the background luminosity of the sky tends to decrease, and the beautiful blue tint we see at sea-level dulls into a deep violet. This is because the number of molecules present to diffuse the light of the Sun is decreasing. As the last traces of diffused light die out, the true blackness of space begins to appear.

It is quite natural that the atmospheric density should fall off with height, as the underlying layers must be compressed by the weight of those above. However, this decrease is not perfectly uniform, and we could not calculate the total amount of air overhead at any particular place simply by knowing the altitude of the place. Nor is the atmosphere uniform in composition or temperature.

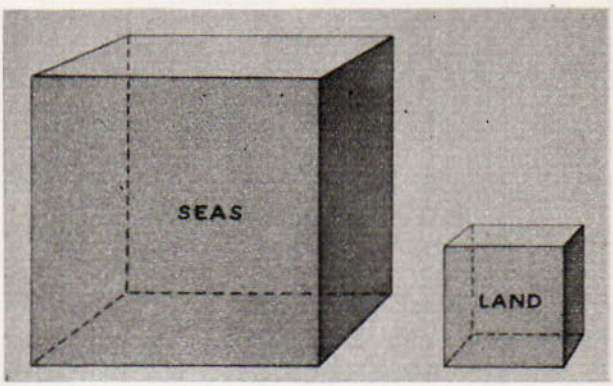


Fig. 184 — RELATIVE VOLUMES OF THE OCEANS AND CONTINENTS

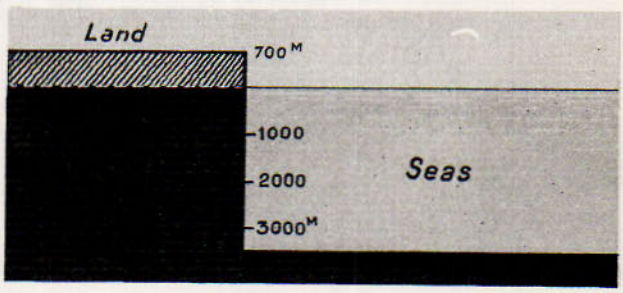


Fig. 185 — COMPARISON BETWEEN THE AVERAGE HEIGHT OF THE LAND AND THE AVERAGE DEPTH OF THE SEA

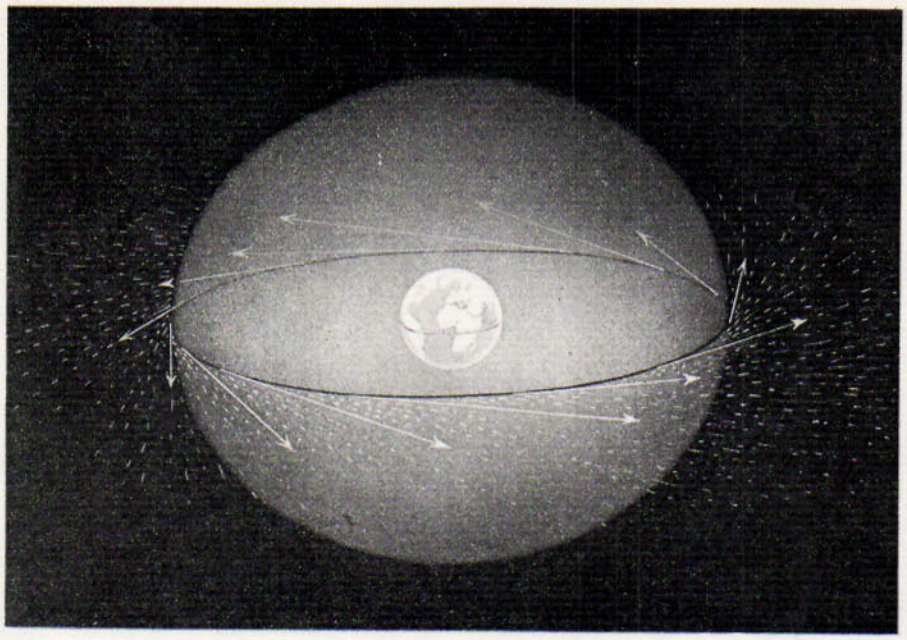


Fig. 186 — THE THEORETICAL LIMIT OF THE ATMOSPHERE IS DETERMINED BY THE MATCHING OF CENTRIFUGAL FORCE AND GRAVITY

Investigation has revealed that it comprises a number of well defined layers, each with its own particular properties. Up to recent years, our investigations were confined to regions readily accessible to our sounding equipment, i.e. hardly more than 19 miles high. For information about greater heights we were forced to rely on the study of aurorae and meteors.

A first approximation of the thickness of the atmosphere can be gained from observing the twilight, discussed earlier. Fig. 187 shows how, for an observer at O, the size of the twilight arc gradually diminishes as the Sun sinks further and further beneath the horizon, whilst on the opposite side of the sky, a region of darkness—the shadow of the Earth — begins to climb up from the horizon. The visibility of this shadow depends largely on the purity of the atmosphere, and if the conditions are favourable, it can be seen in quite a striking manner (*see* Fig. 188).

Then as the general illumination of the sky fades even more, the twilight arc dips beneath the horizon, vanishing completely when the Sun has reached an angle of 18° below the horizon. From this it is possible to calculate trigonometrically that the thickness of atmosphere is 50 miles.

Although this limit is determined by observation, it does not follow that the atmosphere actually stops at this point. First of all, it does not stop abruptly, but gradually fades away to a point where it can have no appreciable effect. This is why night is considered to be total when the twilight has completely disappeared, and the faintest stars can be seen. As yet the sky is not completely dark. The twilight arc, which is clearly visible after sunset, belongs merely to the thickest layers of the atmosphere, and by sensitive optical or photographic methods, it is possible to detect a very faint and diffuse glow before the appearance of the dawn light and after the disappearance of the evening twilight. This glow is best seen when clouds drift across; they can be seen sharply silhouetted

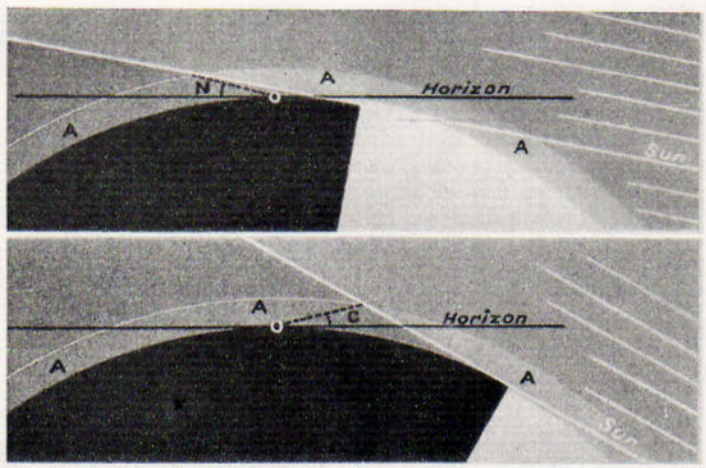


Fig. 187 — FOR AN OBSERVER AT O, THE ILLUMINATED PART OF THE ATMOSPHERE (A) ABOVE HIM DECREASES. At the same time the arc of darkness (N) grows, whilst the twilight arc (C) disappears.

Fig. 188 — THE EARTH'S SHADOW IN THE ATMOSPHERE (THE ONSET OF NIGHT). *(Photo, Ksara Observatory)*

whilst everywhere else they appear only as vague blurs. If this glow is of the same origin as the twilight proper, or something very similar, it can be deduced that the atmosphere still possesses a considerable density at a height of 150 miles. This is supported by other phenomena such as shooting stars (meteors) and aurorae which occur at these heights.

Shooting stars are small particles of material orbiting round the Sun at high speed. When they encounter the Earth, they are rapidly vaporised by friction against the atmosphere, and appear as a sudden streak of light against the stars. Investigation has revealed that they occur at between 50 and 100 miles up. Aurorae, or the Polar Lights as we call them, occur at the same height as meteors but extend much higher, sometimes reaching as much as 500 miles above the surface. Finally, research into the luminosity of the sky has revealed the effects of sunlight at the record altitude of approximately 625 miles. We shall return to this in more detail later.

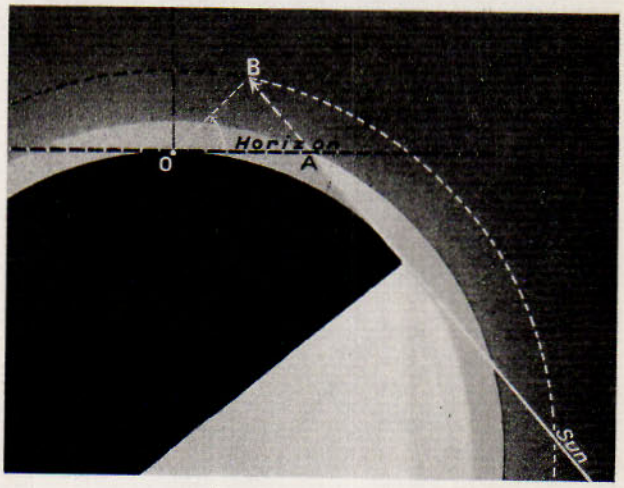


Fig. 189 — FOR AN OBSERVER AT O, THE TWILIGHT ARC PROPER FINISHES AT A. The highest layers of the atmosphere are slightly illuminated as far as B.

Weight and composition of the air. — The general composition of the air was first discovered by Lavoisier

Fig. 190 — A FAINT GLOW AFTER TWILIGHT PROPER HAS DISAPPEARED. The clouds can be seen quite clearly in silhouette

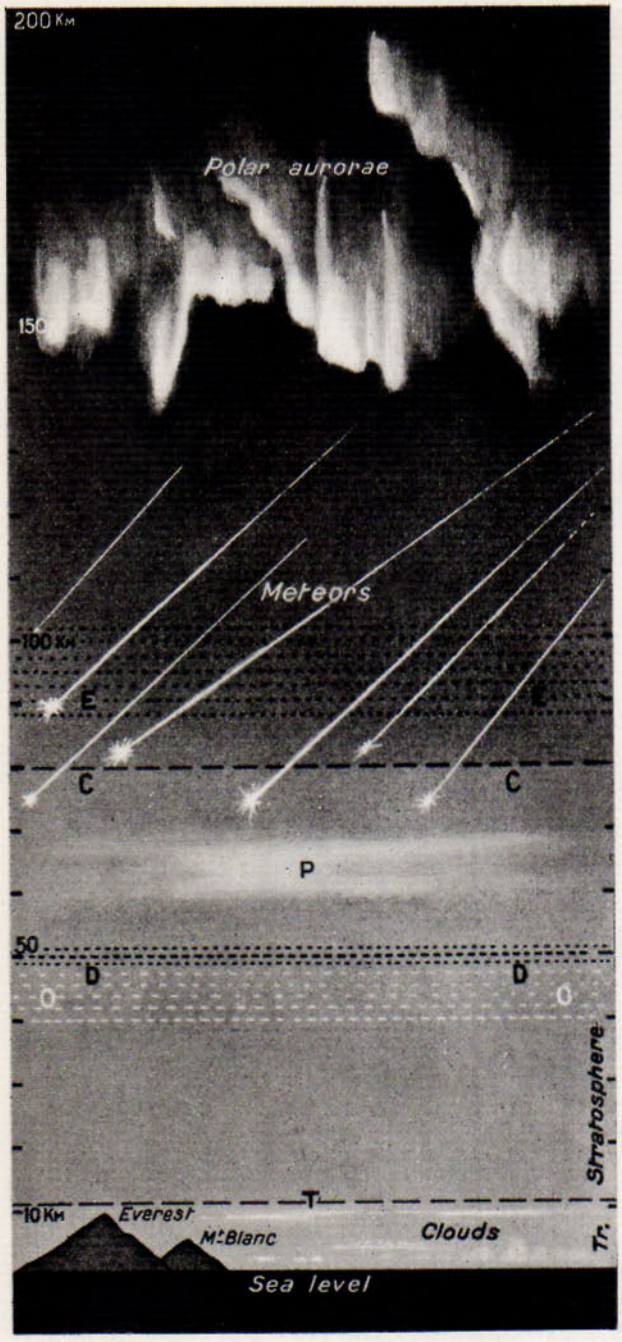


Fig. 191 — EARTH'S LOWER ATMOSPHERE. T, tropopause; O, ozone; D, ionised layer; P, volcanic dust; C, upper limit of twilight; E, Heaviside layer. The polar aurorae are shown at a height of about 150 to 200 km., but they do extend to a height of 900 km. or more.

in 1774. It has been more thoroughly determined by the work of the English physicists Ramsay and Lord Rayleigh

The air that we breathe at the surface of the Earth consists by volume of the following elements:

Nitrogen	0.7803	Hydrogen	0.0001
Oxygen	0.2099	Neon	0.000015
Argon	0.0094	Helium	0.0000015
Carbon Dioxide	0.0003	Krypton, Xenon	traces

Water vapour, it may be noticed, is absent from this list, although it is abundant in the atmosphere. However, its occurrence is localised and variable in quantity. For this reason and because of its special importance, it will be dealt with separately.

The composition of the air varies through the different atmospheric layers, which means that the above proportions of the different gases do not remain constant at different heights. Our information about the higher layers is very meagre, since we are only just beginning to observe them directly, but what we have learned has profoundly modified earlier ideas based purely on theoretical considerations.

First of all, we must consider the amount of air contained in the atmosphere; we shall approach this by way of two barometrical principles.

The weight of one cubic foot of air, under normal pressure at sea-level, is 1.3 oz. If the whole atmosphere were at this pressure under these conditions, it would form a uniform layer over the Earth only 5 miles thick. If the temperature were lowered until the air liquefied, this new sort of ocean would cover the entire surface of the Earth to a depth of about 30 feet.

It is calculated that the total mass of the atmosphere is one millionth that of the globe. This may not sound a great deal, but in absolute terms, the atmosphere still weighs 5,000,000,000,000,000 tons.

Structure of the atmosphere. — As we have said, air becomes thinner with height. Even in the lower parts of the atmosphere, the fall-off is rapid. At sea-level, the barometric pressure is usually 760 millimetres of mercury, whilst at a height of 12 miles, this has dropped to only 41 mm. At about 30 miles, we calculate that it would have dropped to 1.5 mm., and at 60 miles it would be 0.003 mm. Thus at great heights above the surface the atmosphere becomes extraordinarily rarefied, and at 100 miles, it would constitute a more perfect vacuum than we can achieve in the laboratory. At this height, the normal definitions of states of matter as used in physics, become meaningless; for example, a liquid is, so to speak, a continuous substance, with close cohesion between the molecules (although they are not so close as in the solid form). In the same way, what we call a gas is formed of

Fig. 192 — THICKNESS OF THE EXPLORED PART OF THE ATMOSPHERE

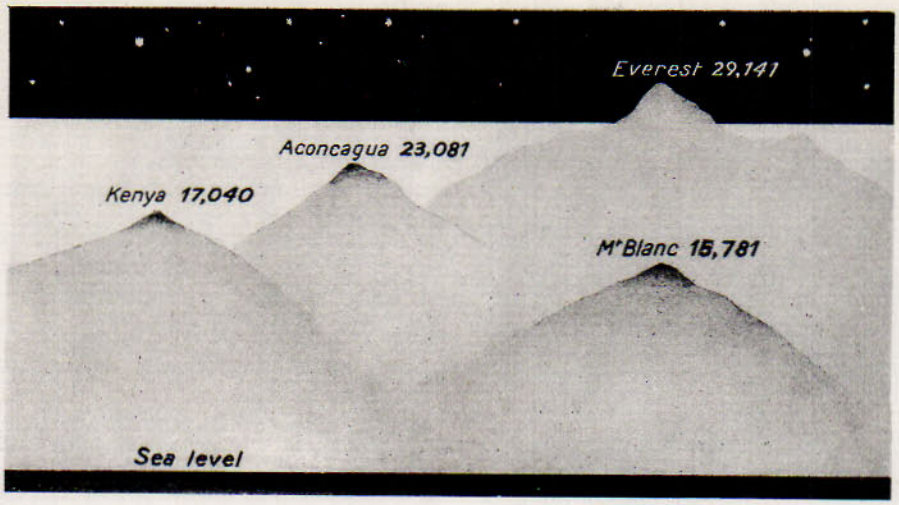


Fig. 193 — THICKNESS OF THE ATMOSPHERE IF IT WERE CONDENSED INTO A UNIFORM LAYER AT SEA-LEVEL DENSITY

individual molecules with no close cohesion between them, flying about in all directions and colliding frequently. In the air we breathe, the mean free path between two molecular collisions is about one ten-thousandth of a milimetre. When the air density decreases, collisions become more infrequent and the mean free path correspondingly longer. This is what occurs at the high altitudes under discussion; there is still matter, but no gas, strictly speaking.

It is rather difficult to form any correct idea of the state of the atmosphere at these great heights; we must rely on theoretical considerations seasoned with a certain amount of observational evidence, and the two are sometimes incompatible.

The law of decreasing density with increasing height would indicate, assuming the atmosphere to be at rest, that the fall-off is closer for light gases than for heavy ones. The percentage of nitrogen would be constant up to about 30 miles, where the amount of oxygen would only be about one-third of its sea-level value, and the proportion of hydrogen would be greater, about 13%. At about 60 miles, nitrogen would be found only in traces, oxygen would be absent, hydrogen would make up about 95% and helium 0.45% of the atmosphere.

As far as we know, the composition of the lower layers remains constant up to a height of 12 miles, but the disappearance of nearly all oxygen and the preponderance of hydrogen in the upper regions does not seem to be confirmed. We shall see that analysis of light from aurorae at a height of several hundred miles, reveals the presence of nitrogen and oxygen; similarly, it is the presence of oxygen that allow us to detect the extension of the atmosphere to a height of 60 miles.

At any rate, it appears that the atmosphere does not fall off uniformly as it would if in a state of equilibrium. Both our direct and our indirect observations reveal a complexity which requires us to distinguish several different regions of the atmosphere and, within these broad divisions, narrower layers all with their own individual characteristics.

Starting with the highest reaches, our main source of information consists of the phenomena already discussed, meteors and aurorae, and the behaviour of electro-magnetic waves, which are reflected by some of the layers under the right conditions.

To explain the reflection of electro-magnetic waves in the atmosphere, we must postulate the presence of charged particles, called 'ions'. Ionisation of the atmosphere has several causes, both within and outside the Earth.

The first of these factors (radioactivity of the crust and certain gases, volcanic activity, etc.) decreases with altitude, that is to say, with distance from the Earth's surface; while the second (cosmic rays, ultra-violet radiation from the Sun, streams of charged particles from solar flares) becomes more important at great heights, until it is so intense as to warrant these regions being called the ionosphere.

Their effect on the propagation of radio waves was recognised right from the very early days of radio. As early as 1902, Kenelly and then Heaviside discovered that electro-magnetic waves can travel right round the Earth, and it was a natural analogy to think of them as behaving like sound-waves, which can pass round an obstacle. This could only be explained on the basis of a charged layer in the atmosphere which could reflect radio waves back to the ground.

It was only a matter of time before this at first bold hypothesis was confirmed experimentally. Two sets of waves were transmitted from one post and received by another at a distance of 60 miles. One set of waves went

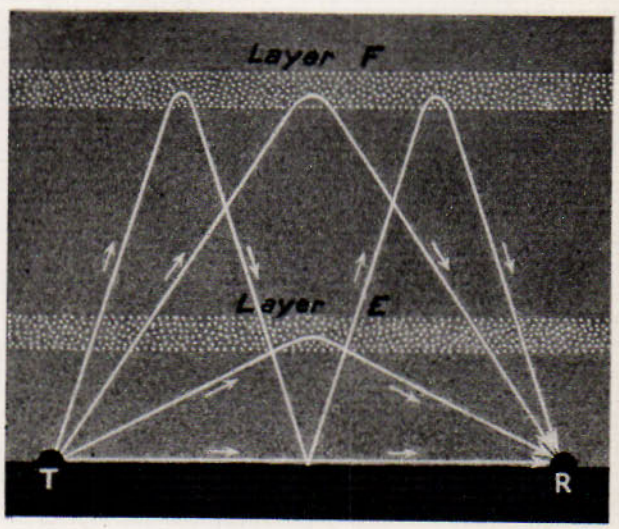


Fig. 194 — REFLECTION OF ELECTROMAGNETIC WAVES FROM IONISED LAYERS IN THE UPPER ATMOSPHERE

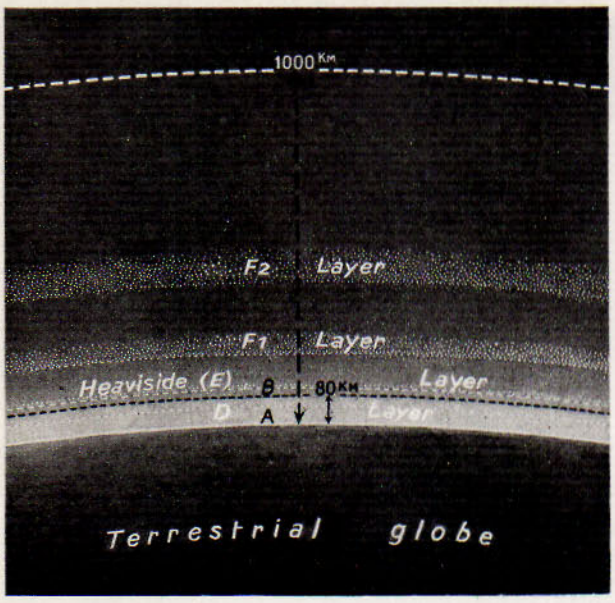


Fig. 195 — THE LAYERS OF THE IONOSPHERE

direct to the receiver, whilst the second was reflected from the ionised layer at some considerable height and arrived slightly later. Nowadays, this can be done more simply: a single station transmits a signal vertically and picks up the returning reflection like an echo. The departure and return are recorded photographically.

This sounding of the atmosphere by means of reflection and refraction of electro-magnetic waves has revealed the existence of a number of ionised layers, differing in thickness, height, and intensity of ionisation. The particular properties of any one layer vary according to the time of day or night.

The principal rôle in the propagation of long waves is played by the 'Heaviside Layer', situated at a height of about 60 miles, and only about a dozen miles thick. This is called the E layer, since there are others at much greater heights; F^1 is about 140 miles high, whilst F^2 is about 300 miles high.

Beneath the Heaviside layer, assuming that it is fairly sharply demarcated, we find a region which marks the end of twilight and shooting stars. Here there seems to be some change in the physical state, perhaps some difference in composition, since the ionosphere begins just above. In the light of modern knowledge, there are reasons for thinking that this region corresponds to the transition zone between the 'low' atmosphere of molecules and the 'high' atmosphere of atoms, where gases such as oxygen and nitrogen are decomposed by the ultraviolet radiation from the Sun.

Nearer the surface there is a zone with which we are unfamiliar, but in which occur a number of phenomena well worthy of our attention.

First, in this region solid particles are able to remain in suspension, this being well attested by the relative density as shown by the existence of twilight. This has been proved by the great volcanic eruptions of the last hundred years, which have violently expelled vast quantites of material in the form of dust into the atmosphere. These dust clouds have persisted for very long periods on account of the minuteness of the particles of which they were composed, and have caused remarkable illumination effects in the sky. This occurred after the cataclysms of Krakatoa in 1883 and Mont Pelée in 1901. The dust is calculated to have reached heights of 40 miles. Strange clouds have also been seen at great heights, glowing in the night and apparently swept along by violent currents of air.

At a height of about 30 miles, there is another layer, the D layer, which appears to be slightly ionised in the daytime but which clears during the night. Between an altitude of 12 and 30 miles we find a region particularly rich in ozone, which is important in many respects.

We finally come to the limits of exploration by means of sounding balloons, and here the task must be taken over by rockets. However, human beings have not tired of trying to explore the atmosphere in person; following such men as Professor Piccard and Stevens, another attempt was made by Major David Simons on 19 August 1957 in the U. S. A. when a new record altitude of about 100,000 feet was reached.

This direct kind of exploration has considerably improved our knowledge of the composition of the atmosphere near the ground. From 1883 onwards Teisserenc de Bort showed by the use of sounding balloons that there are two principal layers, each with its own characteristics; they are the stratosphere and the troposphere, separated by the tropopause.

Generally speaking, the stratosphere, the upper limit of which has not yet been discovered, is homogeneous in structure and almost constant in temperature.

Immediately beneath it is the tropopause, but this layer is not at a uniform height above the ground; it does not follow the curvature of the Earth. At the equator it is at a height of about 10½ miles, whilst at the poles it is about 5½ miles high. At intermediate latitudes it is about 7 miles high. This, however, is only an average figure; under certain conditions variations of several miles may occur.

The troposphere is the layer in contact with the Earth's surface. Neglecting irregularities such as mountains, the average height of the troposphere is 7 miles. In spite of this, the troposphere alone comprises more than three-quarters of the Earth's total atmosphere, and probably all the suspended water vapour. It may be divided into two main zones.

The lower part of the troposphere, lying between the surface and a height of 2 miles, is the zone where meteorological phenomena occur: depressions, anticyclones, hurricanes etc. Its constant turbulence ensures a continual mixing of the atmospheric gases so that their proportions always remain the same. It is here that nearly all the water vapour is found. The upper half is the larger; it is far less disturbed than the lower, and the movements of air which take place are of a different nature.

The temperature of the atmosphere. — As in the case of the density, exact knowledge about the temperature of the atmosphere can only be obtained by direct observation, and beyond this we must rely on what information can be gleaned from phenomena occurring at high altitudes. Unfortunately there is insufficient space to go fully into the theoretical physics involved, so we must be content with a summary of the known facts.

Everyday experience tells us very quickly that the temperature decreases with height: anyone who has taken a trip up into the mountains can vouch for this. Whilst it may be very hot in the plains and valleys, neighbouring summits will be covered in snow from a clearly defined level, and the existence of this 'snow-line' depends on the particular meteorological characteristics of the place, also on its latitude and geographical position. For example, the snow-line in Europe varies from practically sea-level in the Arctic to about 5,000 feet in Norway and 8,000 feet in the Alps. In the Pyrenees it is slightly lower. In the tropics it is at more than 16,000 feet for the

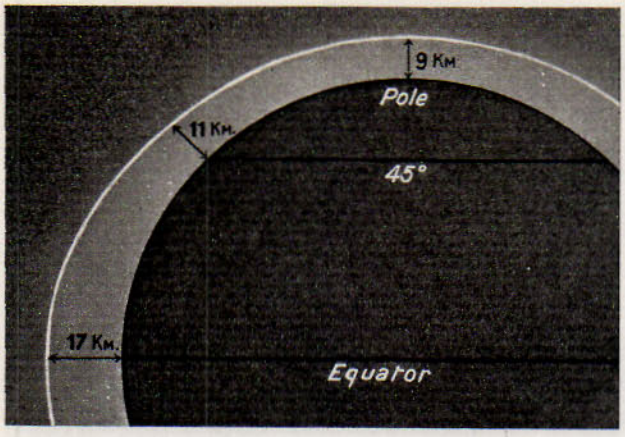


Fig. 196 — HEIGHT OF THE TROPOPAUSE IN VARIOUS PARTS OF THE WORLD

Himalayas, and about 15,500 feet for the Andes.

Meteorology is a specialised branch of learning, and lies outside the scope of this book. We shall therefore confine ourselves to the essential details concerned with the physical activity of a planet.

First of all, the air temperature undergoes considerable diurnal variations owing to the differences in the solar radiation received. This depends on the time of day, the season of the year, etc. The heat in any particular area also varies according to the geographical position and the topography; over land, which is very irregular as a rule, there are liable to be more variations than over a large expanse of water.

This perpetual interaction between the air and the surface of the ground is the basic cause of the general air circulation of the globe, though there are smaller, more localised disturbances superimposed. There is a constant state of turbulence in the lower atmosphere, the home of the phenomenon we call 'weather'. Temperature decreases with altitude because, up to a certain height, the atmosphere radiates more energy into space than it receives from the Sun. Below 12,000 or 13,000 feet this decrease is not uniform and inversions are not infrequent. (An inversion occurs where a warm layer of air overlies a cold one and usually happens on fine calm nights.) This gives some indication of the complexity of meteorological phenomena.

Ignoring these irregularities, the decrease in temperature in middle latitudes up to a height of about 13,000 feet is just under half a degree per 300 feet. Above this the rate becomes more rapid and more uniform, at about 0.70°C every 300 feet.

Above about 29,000 feet, the upper level of the

Fig. 197 — THE TRAIL LEFT ACROSS PHOTOGRAPHIC PLATE BY THE FIRST ARTIFICIAL SATELLITE
(Courtesy British Interplanetary Society)

Fig. 198 — THE TEMPERATURE OF THE ATMOSPHERE AS A FUNCTION OF ALTITUDE. The snowline in a mountain massif.

troposphere, the fall in temperature decreases and reaches its minimum at an altitude which varies with the season. From this point on, the temperature is practically constant. However, it is not the same above different parts of the globe, for we have arrived at the base of the stratosphere. At the equator, where the stratosphere begins at a height of 10½ miles, the temperature is −86°C; at the poles (5 to 6 miles), it is −52°C, whilst at latitude 45° (6½ miles), it is −56°C. In contrast with conditions on the surface, it is much colder, at a given altitude, over the tropics than over the polar regions.

It is quite possible that this trend does not continue much above the limits reached by sounding balloons. Recent research with high-altitude rockets shows that at very great heights the temperature starts to rise again.

It has also been observed that, on occasion, sounds can travel enormous distances, far exceeding their normal range. It seems that, as in the case of radio waves, sounds can be reflected from one layer to another, where, in addition, their speed increases above normal. This may mean the existence of temperatures of several hundred degrees in the upper atmosphere. It must be emphasised, however, that the idea of 'temperature' loses its connotation of 'heat' in the rarefied air of these regions, and it should be understood simply as an expression of the speed of thermal agitation of the molecules.

Clouds. — We must spend a few minutes studying the clouds if only because they formed an obstacle before which astronomers had to confess themselves defeated, at least until the invention of the radio-telescope. An observer in space, moreover, would have the greatest difficulty in picking out the familiar shapes of the continents through the shifting veil of clouds and dust which covers the Earth. In this connection, a study of the disturbances in our own atmosphere should be useful when we come to interpret our observations of other planets.

The ever-changing appearance of the clouds and the incomparably beautiful effects of light and shade that they produce result from the perpetual turbulence of the air near the surface of the ground. Here we shall content ourselves with stating the underlying principles of cloud formation.

In the atmosphere there exist large quantities of water vapour; this gas is quite invisible and results from evaporation of the seas, rivers and lakes of the world. The amount of water vapour present in a given volume of air depends on the temperature. When this falls below the point at which saturation occurs, condensation in the form of ice-crystals or water droplets takes place around tiny dust particles and charged particles suspended in the atmosphere which act as nuclei. Clouds are composed of such droplets and crystals.

Despite their transient nature and their sometimes sharp, sometimes blurred outlines, clouds are not the result of chance. A little study will show that they fall into a number of quite definite types, determined by the temperature and humidity conditions prevailing when they are in process of condensation from water vapour.

Masses of air cool down by three distinct processes: radiation, mixing and expansion, which may be summed up in the following way. Radiation, e. g. the emission of heat during a clear night, will cool down the Earth, and therefore the air in contact with it. The water vapour will be deposited as mist or dew. In the same way, a mass of air may cool down by radiating its heat into space, but this process is not very important because of the low efficiency of air as a radiator.

The mixing of two air masses containing water vapour at different pressures results in condensation if the process of mixing raises the level of saturation above the mean of the two masses. This form of condensation produces only mist and light cloud, and the actual precipitation is very slight.

Expansion is by far the most important factor. When the pressure on a gas is lowered, the gas will naturally expand, and this causes a drop in temperature. Conversely, when a gas is compressed, it becomes heated, as we can feel for ourselves when pumping up a bicycle tyre. Thus, when air in contact with the ground is forced to rise, the pressure drops, the air expands and the temperature falls. This fall in temperature occurs at a rate of approximately 1° per 300 feet for dry air, and 1° per 600 feet for saturated air. This is the process which produces very large masses of cloud, and is the main source of precipitaiton. This mechanism depends on a number of factors, the general air circulation of the planet, the geographical location, the configuration of the ground, all being very closely connected.

Expansion can take place either thermally or mechanically. In the first case, there may be a general upward movement of air over a large area of heated ground. Once the process has started, it tends to maintain itself. In the

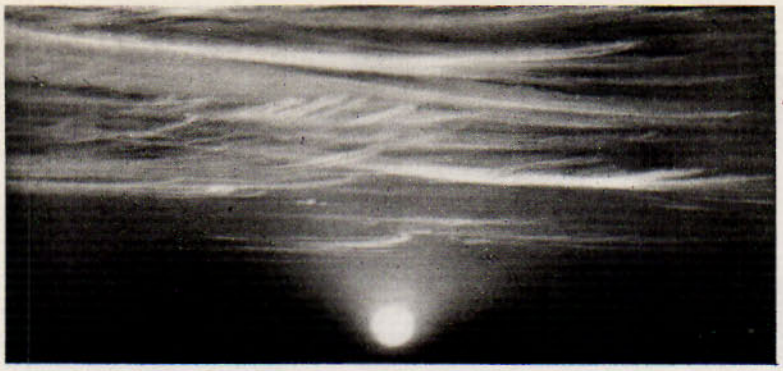


Fig. 199 — CIRRUS

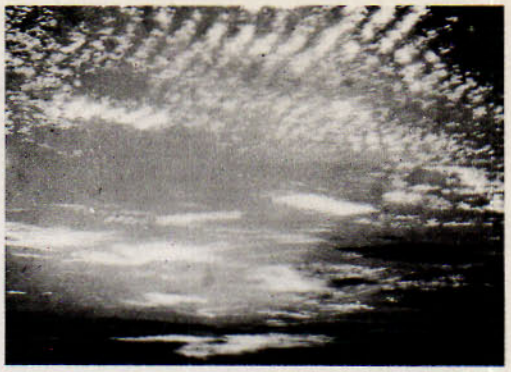


Fig. 200 — CIRRO-CUMULUS

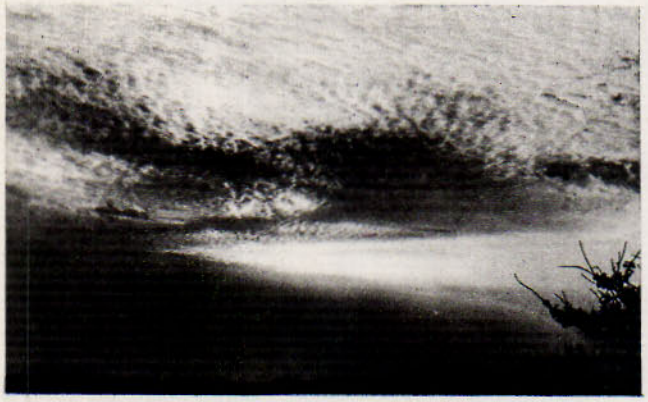


Fig. 201 — ALTO-CUMULUS

Fig. 202 — ALTO-STRATUS

Fig. 203 — CUMULO-NIMBUS

Fig. 204 — CUMULUS

Fig. 205 — NIMBO-STRATUS

Fig. 206 — STRATUS

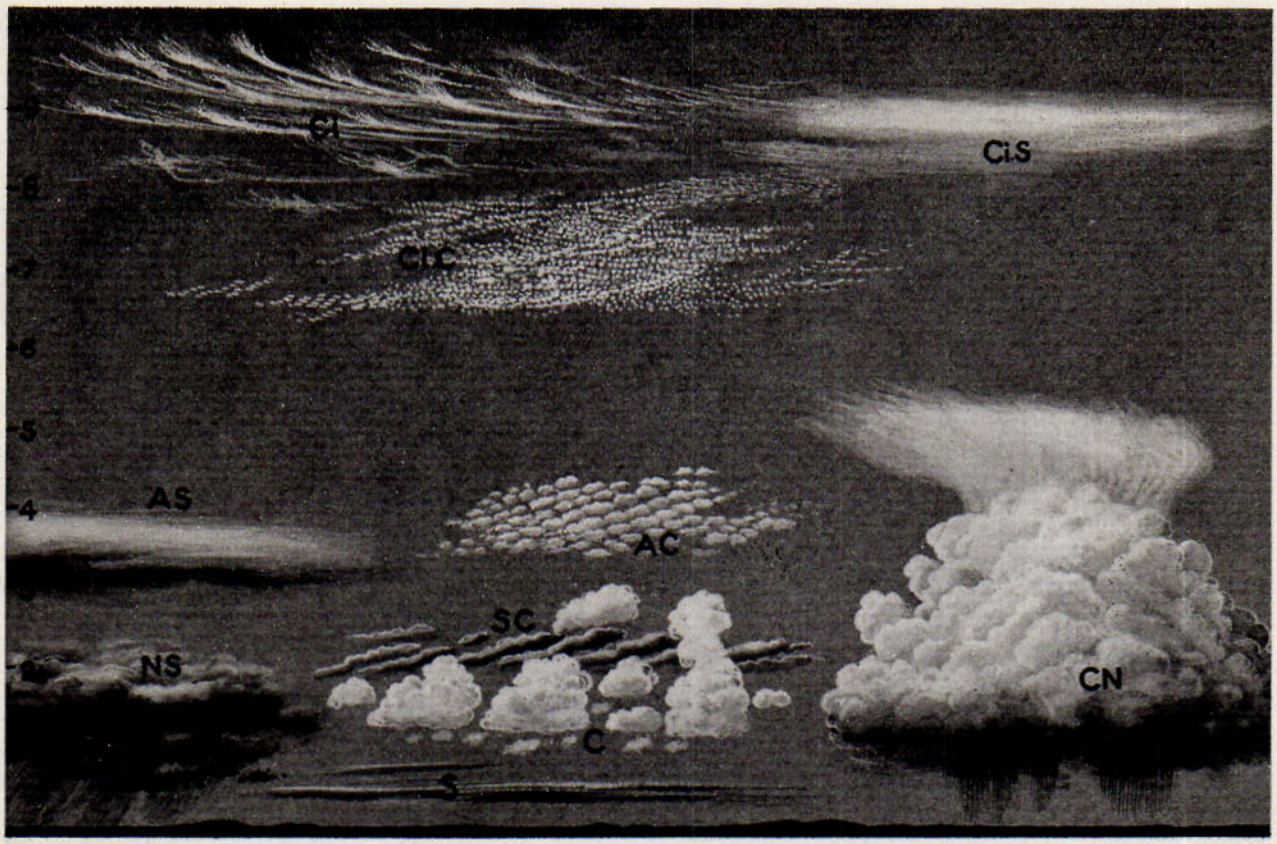


Fig. 207 — AVERAGE HEIGHT OF DIFFERENT TYPES OF CLOUD. S, stratus; N, nimbus; C, cumulus, N-S, nimbo-stratus; S-C, strato-cumulus; C-N, cumulo-nimbus; A-S, alto-stratus; A-C, alto-cumulus; Ci-C, cirro-cumulus; Ci-S, cirro-stratus; Ci, cirrus.

second case the air is forced upwards by having to pass over an obstacle, such as a range of mountains or a plateau. Warm air will sweep up the mountain sides, and at a certain height precipitation will begin, forming those masses of thick cloud which often seem to gather round mountain tops. Generally speaking, high places receive more precipitation than low ones, in the form of either snow or rain.

The height at which any particular type of cloud forms therefore depends on the prevailing conditions, which vary from season to season and from one part of the Earth to another. The most typical example of this is cirrus cloud, the highest of all, which may on occasion occur at less than half its normal altitude, and in polar regions can come right down to ground level. On the other hand, strong updraughts may form clouds at heights greater than those of normal cirrus cloud. This clearly shows the complexity of the subject.

The different types of cloud found in temperate regions are shown in Fig. 207. To this we may add a few words about their individual characteristics.

The products of condensation are, as we have said, of two sorts. At temperatures below 0°C, precipitation is in the form of minute hexagonal ice crystals. Otherwise it consists of water droplets, which when formed are about 0.000001 to 0.001 millimetre in diameter. They are therefore completely invisible to the naked eye, but as they cool down, they grow to a diameter of about 0.01 millimetre, causing a mist or cloud.

These crystals and water droplets begin to fall under gravity as soon as they are formed, but because of their minuteness, air resistance becomes quite an important factor, since the weight of a spherical body is proportional to the cube of its radius, while the surface area depends on the square of the radius. The relation of surface area to weight therefore increases when the radius decreases. This means that these particles will fall very slowly, at only a few feet per second, and if they are caught in an ascending air current, they may well start to rise.

We are of course here dealing with what constitutes the visible mass of a cloud suspended in the atmosphere.

Fig. 208 — FORMATION OF CLOUD; UP-DRAUGHTS

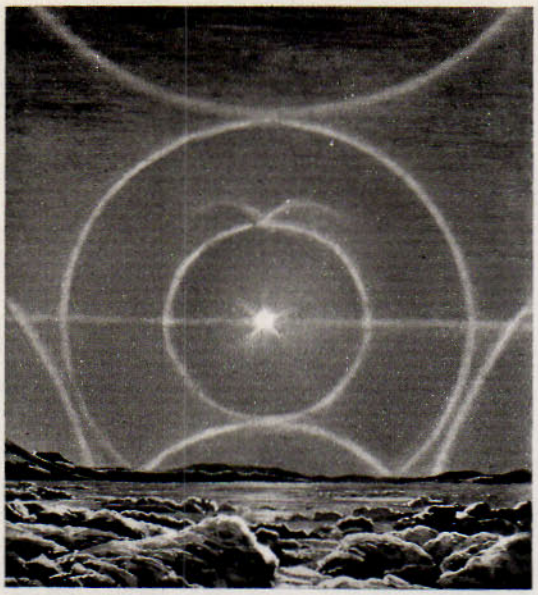


Fig. 209 — COMPLETE SOLAR HALO

When the diameter of the water droplets reaches more than a tenth of a millimetre, air resistance and updraughts are not enough to prevent them from falling. This marks the beginning of real precipitation, the mechanism of which is even more complex than was once thought. Raindrops, which usually have a diameter of between 1 and 5 millimetres, are at least a thousand times larger than the primitive droplets constituting the cloud, and cannot be explained by simple condensation. What happens is that some of the particles in the cloud must coalesce, either through collision between droplets or indirectly by the growth of some at the expense of others; in any case, it appears certain that electricity plays a very important part. Rain, however, does not fall from every type of cloud. Continuous precipitation is essentially due to alto-stratus and nimbo-stratus; on the other hand, violent downpours usually come from cumulo-nimbus.

A cloud cannot be considered simply as a mass of droplets carried along by the wind. It is the visible manifestation of a disturbance, growing independently of the stream of air in which it exists. In other words, once a cloud starts to form, it affects the surrounding air and tends to grow rapidly.

Clouds are of characteristic shape, structure and appearance, depending on their mode of formation, and by this they are easily distinguishable. The delicate, complex filaments of cirrus, and their extension as a cirro-stratus veil (cirro-cumulus, on the other hand, being characteristically nodulated in structure), are composed of tiny crystals, since they are at an altitude where the temperature is extremely low. The hazy clouds of crystals are not very thick, and the light which passes through them is refracted as if by myriads of tiny prisms, often producing a magnificent halo, which is a large, slightly iridescent circle centred on the Sun or Moon. The phenomenon usually appears as a single circle, but sometimes it is much more complex with a number of different circles, some concentric, some making tangents, accompanied by patches of shimmering light called parahelia, or sun-dogs, which are the refracted images of the Sun. This beautiful sight at once gives us a clue to the state of the atmosphere at the time.

A similar phenomenon, often confused with the halo proper, is the corona (which should not be confused with the Sun's corona as seen in an eclipse). This takes the form of a series of bright circles, whitish or coloured, round the Sun or Moon. Whilst, however, a halo is an immense circle 23° in radius, a corona is only a few degrees across, and very often is no more than a whitish patch around the body in question. The two phenomena have different causes: a corona is not the result of refraction, but of the reflection and diffraction of light by tiny spherical droplets of water forming the lower clouds. From this it is easy to tell at what height the clouds concerned are located.

Cloud systems: cloud distribution. — Clouds are not distributed at random through the atmosphere, any more than they are formed by chance. They are produced by great meteorological disturbances which arise over various parts of the Earth, and generally speaking, they are associated in cloud-systems. A study of weather maps compiled from the observations of all the meteorological services of the world has shown that places where it is raining at any particular moment are engulfed in much larger areas of cloudy sky. On the fringes of these areas, in the front and of either side, are belts of cloudy sky, whilst behind them there is usually a vast area of very varied weather. It may be wondered why the terms 'in front' and 'behind' are used. The reason is that this whole mass moves as a system in a definite direction, and this system may be very large in extent — often much greater than the British Isles, for example.

The structure of a cloud system is as follows: at the head is a front, containing very high cloud such as cirrus, cirro-stratus, or cirro-cumulus, distributed along a rather narrow belt. The central area contains medium high to low clouds, cirro-stratus, nimbo-stratus, forming a thick, rain-producing blanket. At the edges are more high clouds, mainly alto-cumulus, fringed with cirrus. In the rear is a zone of cloud and clear patches, as large as the rest put together. Here the principal clouds are usually the tall cumulo-nimbus, which give off sudden, often violent downpours of rain, snow or hail. There are often sudden gusts of wind and squalls, occasionally accompanied by electrical effects.

Broadly speaking, these systems usually retain their relative position when moving. They are frequently separated by intervals of very clear sky; sometimes, too, they follow in an uninterrupted stream, even overlapping on occasion. Others are less clearly defined, and are liable to rapid change: these are the storm systems.

Fig. 210 — LUNAR CORONA

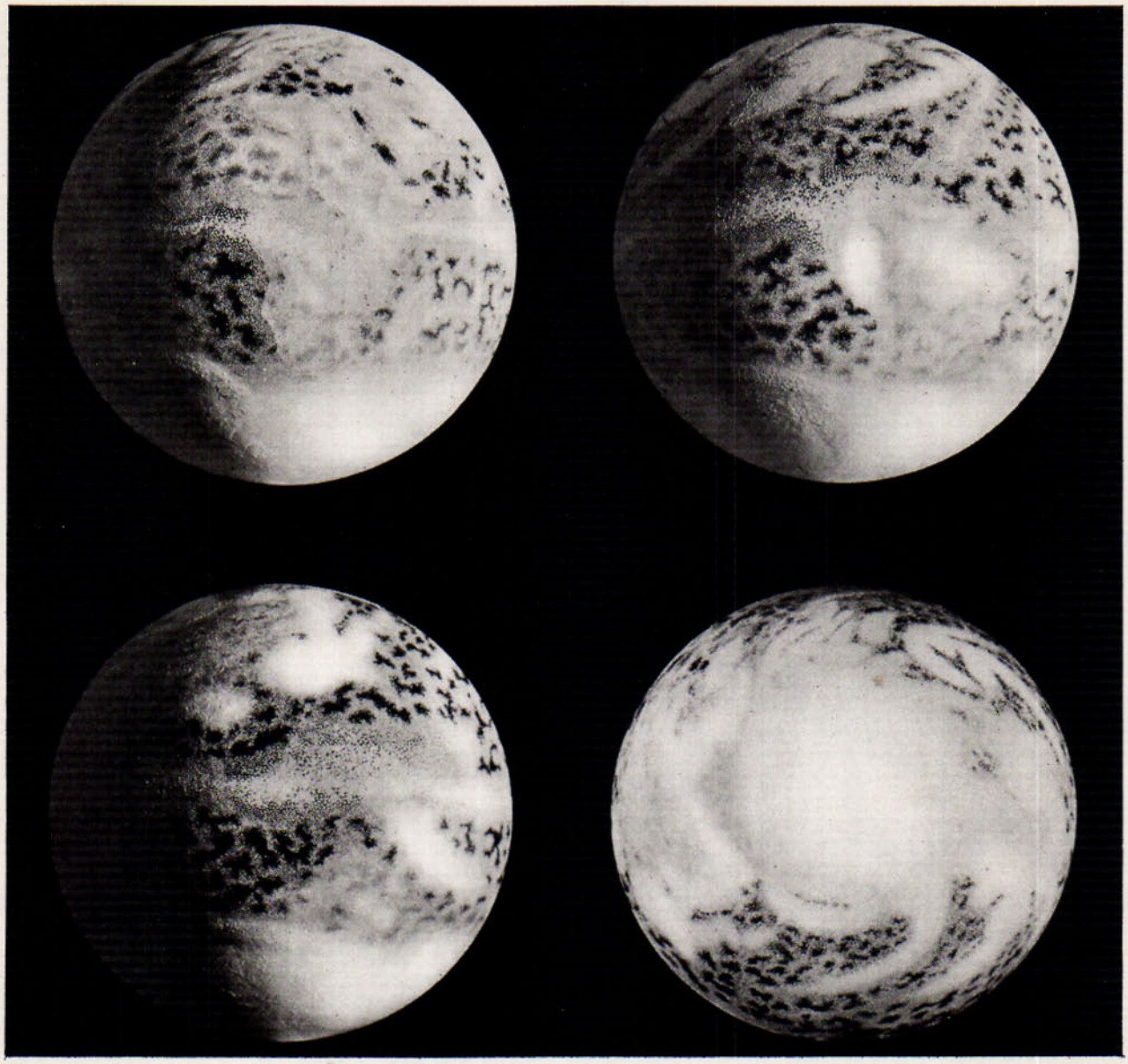


Fig. 211 — THE EARTH SEEN FROM SPACE; THE WELL-KNOWN FEATURES ARE HARDLY VISIBLE THROUGH THE ATMOSPHERIC VEIL. From left to right: (above) Africa, South America; (below) Pacific Ocean, North Pole. Compiled from international meteorological reports, spring equinox, 1935.

When we study the world as a whole from the meteorological point of view, (and this is becoming increasingly possible thanks to the expansion of the network of 'met' stations), we begin to realise that the atmosphere is so charged with cloud that an observer in space would have the greatest difficulty in picking out the well-known pattern of oceans and continents. Our planet would seem curiously mottled owing to the superimposition of all the different cloud systems over the surface features, and the whole picture would be in constant turmoil. Although this study of the state of the atmosphere is both interesting and instructive, it is really outside the bounds of astronomy, except that a knowledge of what goes on here on Earth may help us to interpret what we see on other worlds.

The colour of the atmosphere. — We all admire the beautiful blue of a clear sky or the superb contrast between the brilliant white of the clouds and the azure between them. This colour, however, is not inherent in the atmosphere: it is an optical effect.

When light encounters a cloud of small particles, it is diffused. This may be illustrated by a familiar example. When sunlight enters a darkened room through a gap in the blinds, it will appear as a brilliant ray of light. This is because there are tiny particles of dust suspended in the air, and these, although normally invisible, catch and diffuse any strong light shining on them. The same effect can be seen in the sky when broken clouds hide the Sun; the glorious shafts of light then formed are shown in Fig. 212. Thus we can see that diffusion is occurring through all sorts of impurities in the atmosphere: dust, vapour, even molecules of gas. This explains all the delightfully varied shades of colour to be seen at different times of the day.

White light is a mixture of all the colours of the rainbow or spectrum, which, as Sir Isaac Newton proved, can be separated by a prism. In order, they are red, orange,

Fig. 212 — DIFFUSION OF LIGHT BY VAPOUR AND DUST IN THE ATMOSPHERE

yellow, green, blue, indigo, violet, and they are all to be found in the solar spectrum, to which we shall return in Chapter 20. For the moment we may consider light as made up of an infinite number of simple radiations; these radiations can be likened to a series of waves, diminishing in length from one end of the spectrum to the other, red light having the longest and violet having the shortest wavelength.

Diffusion acts in a very unequal way over this spectrum and Lord Rayleigh established that its effect is inversely proportional to the fourth power of the wavelength. Violet light is therefore diffused approximately 16 times more strongly than red.

At this point we must stop to consider the dimensions of the particles in question. The smaller they are, the greater will be the proportion of light of short wavelength contained in the diffused light, thus achieving the ideal conditions mentioned. This is exactly what happens in the air: the size of the molecules is infinitesimal (the nitrogen molecule is about 0.000003 millimetre in diameter) and in these conditions, the preponderance of radiation at the blue end of the spectrum will give this colour to the diffused light.

But this blue, characteristic as it may be (giving rise to the expression 'sky-blue'), is not an invariable quantity, or an absolute standard, but ranges from very dark to very pale. At sunrise and sunset, however, the sky often takes on a wonderful fiery colour, which at first glance would seem to contradict the above theory; but this does explain all the effects if we make proper allowance for all possible conditions obtaining in the atmosphere.

Once more, it must be remembered that the size of the particles is important. Impurities such as dust and water droplets are much larger than the molecules of gas. In the lower parts of the atmosphere, where most of these impurities are concentrated, the colour of the sky is quite different from what it is at height where no impurities are present. A simple experiment will demonstrate this principle. Examine the smoke from a cigarette before and after it has been inhaled. At first the smoke will be plainly bluish in colour, since it is made up of very minute particles; after it has been inhaled and then blown out through the mouth, it will have taken on a yellowish tinge, because it now contains water-droplets which have condensed round the particles and which are much more efficient diffusers of light.

This is why the colour of the sky when seen from the top of a mountain is much more intense and dark blue than when seen from sea-level. In the first place, there is an almost complete absence of water vapour and, more important, dust. Secondly, we are above the thickest portions of the atmosphere, which naturally have the most powerful diffusing effect. The colour of the sky varies from horizon to zenith, since there is a different thickness of air in these two direction. The horizon always appears lighter, more whitish or even slightly reddish, according to the state of the atmosphere. All this, incidentally, also intensifies the illusion of being under a great dome, the 'celestial vault'.

Generally speaking, these same conditions, shown diagrammatically in Fig. 213, operate unequally through the day. When the Sun is near the horizon, its light must traverse the thickest part of the lower, dust-laden atmosphere; this means that in the direction of the Sun the sky will appear bright through the preponderance of yellow and red wavelengths in the light, and the Sun will become redder and duller as it nears the horizon. The same thing can be seen in the case of the Moon and the planets when viewed through a telescope. This colouring occurs when the white light is deprived of the blue and violet wavelengths, which are diffused in all directions.

The intensity of all these beautiful colours is nevertheless extremely variable, as may easily be seen. The percentage of water vapour present in the atmosphere is very important in this connection. We have already seen how condensation forms clouds and mists, which are made up of very small spheres of liquid. These reflect and refract all wavelengths of light so that the same proportions remain in the diffused light. Thus when there is even a slight mist about, the sky loses its beautiful blue colour and takes on a milky appearance.

The behaviour of light rays passing through our dust-

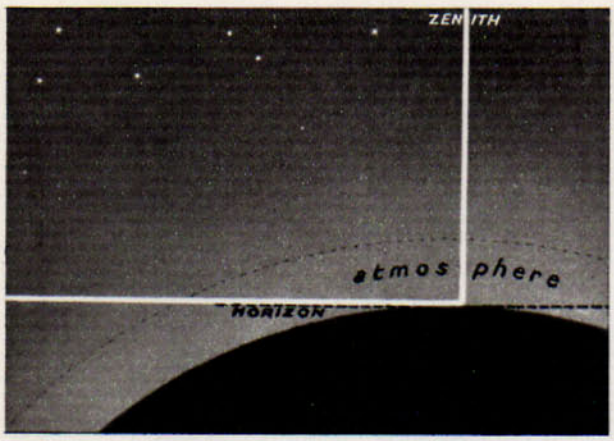


Fig. 213 — ABSORPTION OF LIGHT IN THE ATMOSPHERE. The amount of absorption is maximum near the horizon, where the light has to traverse a greater thickness of air than at the zenith.

and water vapour-laden atmosphere is very important when we are studying the stars and planets. We shall return to this topic in a later section.

Absorption and refraction of light in the atmosphere. — The atmosphere not only diffuses the Sun's rays to give us the general illumination we call daylight, and decks the landscape with beautiful mists, but it also acts as a sort of screen. This screen has several important functions. It behaves rather like the glass of a greenhouse, preventing the heat which has been stored up during the day from radiating into space during the night. Moreover, it softens the effect of the Sun's radiation on living organisms, as we shall see in more detail in Chapter 11.

But, astronomically speaking, the atmosphere is a most formidable obstacle to the observation of the outside universe. It dims the light of the stars and blurs telescopic images, quite apart from causing other distortions.

The effect of the spherical nature of the atmosphere is more pronounced, as we have seen, when looking towards the horizon, because the thickness of air that the light has to penetrate is greater. To demonstrate how effective the light-absorption of the atmosphere really is, we have only to observe the Sun on a summer's day: at noon, when it is high in the sky, we can hardly bear to look at it directly even for an instant, but at sunset it can be stared at without injury to the eyes. Every amateur photographer knows, usually to his cost, how greatly the illumination diminishes in the evening. Indeed, the photographic effect is more pronounced than the visual, since it is the blue end of the spectrum, to which the photographic emulsion is most sensitive, which is absorbed the most strongly, the human eye being adapted to the red and yellow wavebands, where the maximum solar radiation occurs. This it can be understood what an obstacle the atmosphere is to astronomers when they need to photograph some object low down in the sky, near the horizon.

There is still another property of the atmosphere which is the source of even more serious problems.

When passing through any transparent medium (other than space) the direction of a ray of light is changed: the light is said to be refracted. This means that when we see a star lying in a certain direction, its apparent position, owing to refraction, does not correspond to its true position in space. It will be remembred that we have already discussed this in connection with sunrise and sunset. Fig. 216 shows the effect diagrammatically. The atmosphere as a whole is made up of a number of layers of decreasing density, and a ray of light coming in from a distant star must pass through them. In the first layer, refraction will direct it along AB into the second, where it will be refracted more strongly, along BC, and so on. It will finally reach the observer along DE, which will indicate its apparent direction. Thus while the star is really at S, the observer will think it is at S′.

Bearing this principle in mind, we must also remember that the atmosphere, simplified in Fig. 216, does not fall into clearly defined layers, but merely decreases gradually in density with height. This means that the path of a ray of light is really a curve, not a segmented as shown in the figure. The final result is the same, however; the observer sees the star in the direction OS′, which makes a tangent with OS.

When a star is directly overhead at the zenith, its light is coming in at the vertical and refraction is nil; but as the star moves away from this position, its light must pass through an increasing thickness of air, and the effects of refraction increase likewise, reaching their maximum at the horizon.

Refraction will therefore introduce errors into all measurements of stellar positions. In order to obviate them, the laws of refraction have been formulated after experimental study, and these enable the observed positions to be corrected. Let us imagine two stars, A and B, at different angles above the horizon at the time of observation (Fig. 218). Their true positions on the celestial sphere are separated by the angular distance AB. Owing to refraction, however, we see them at A′ and B′ respectively, the shift in position of the one nearer the horizon being greater than that of the other: thus they

Apparent height (h′)	Refraction (R)	Apparent height (h′)	Refraction (R)	Apparent height (h′)	Refraction (R)	Apparent height (h′)	Refraction (R)
0° 0′	36′36″	7°30′	7′11″	34°	1′29″	63°	31″
10	34 18	8 0	6 47	35	1 26	64	29
20	32 14	30	6 25	36	1 23	65	28
30	30 21	9 0	6 4	37	1 20	66	27
40	28 38	30	5 46	38	1 17	67	26
50	27 3	10 0	5 30	39	1 14	68	24
1 0	25 37	11	5 1	40	1 12	69	23
10	24 18	12	4 36	41	1 9	70	22
20	23 5	13	4 15	42	1 7	71	21
30	21 58	14	3 57	43	1 4	72	20
40	20 56	15	3 41	44	1 2	73	18
50	19 59	16	3 27	45	1 0	74	17
2 0	19 7	17	3 14	46	0 58	75	16
10	18 18	18	3 3	47	0 56	76	15
20	17 32	19	2 53	48	0 54	77	14
30	16 50	20	2 44	49	0 52	78	13
40	16 10	21	2 35	50	0 50	79	12
50	15 33	22	2 28	51	0 49	80	11
3 0	14 59	23	2 21	52	0 47	81	10
20	13 56	24	2 14	53	0 45	82	8
40	13 1	25	2 8	54	0 44	83	7
4 0	12 12	26	2 3	55	0 42	84	6
20	11 28	27	1 57	56	0 41	85	5
40	10 49	28	1 53	57	0 39	86	4
5 0	10 13	29	1 48	58	0 37	87	3
30	9 27	30	1 44	59	0 36	88	2
6 0	8 46	31	1 40	60	0 35	89	1
30	8 11	32	1 36	61	0 33	90	0
7 0	7 39	33	1 32	62	0 32		

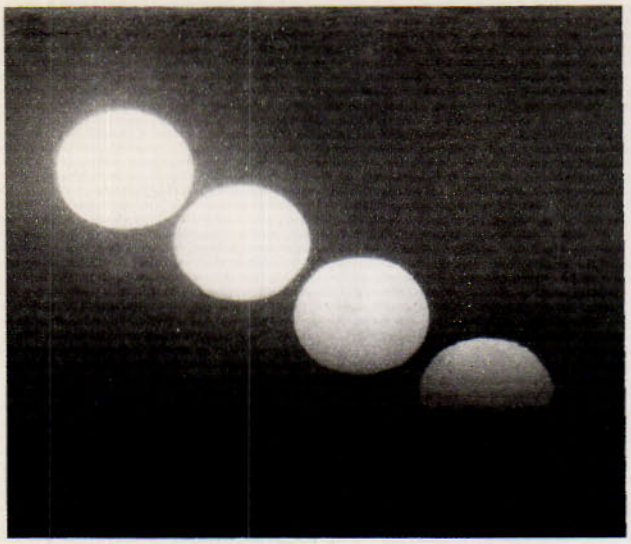


Fig. 214 — DECREASE IN THE BRIGHTNESS OF THE SETTING SUN OWING TO ABSORPTION

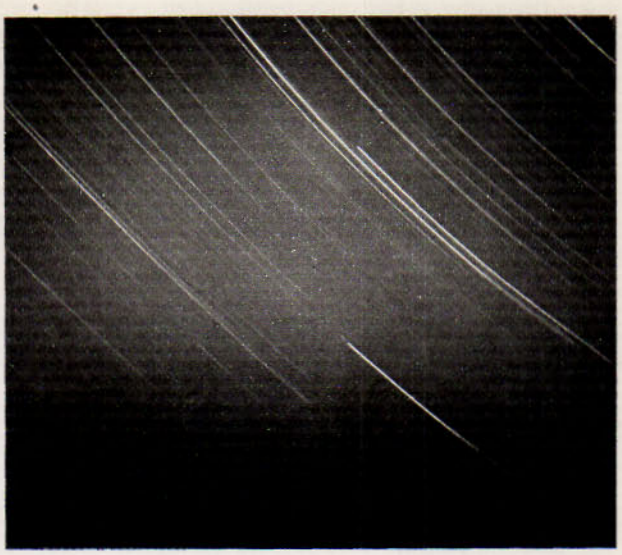


Fig. 215 — EXTINCTION OF STARS NEAR THE HORIZON

will not seem to be as far apart as they ought to be. As the diurnal movement of the Earth changes the position of the stars with respect to the horizon, the same observations made at a different time would give a different value for the *apparent* separation of the two stars. Hence the need for a table of corrections.

Such a table is given here. Refraction is calculated for sea-level at latitude 45°, with a barometric pressure of 760 mm. of mercury at 0°C. Supplementary tables are usually given in astronomical handbooks in order to correct for variation in temperature and pressure. In spite of this, one reservation must be made: the amount of refraction suffered by a ray of light passing through a mass of air depends on the densities of the air layers in question, that is to say, the composition, temperature and humidity. The relation between these is constantly varying, sometimes quite abnormally. Moreover, the different layers are not usually arranged with any regularity.

Nearly all these influences are confined to the lower portions of the Earth's atmosphere, immediately above the surface. Clearly defined layers of air frequently occur in this region, producing a stratification of the atmosphere such as we discussed at the beginning of this section. Rays of light then have to pass through many irregularities before reaching the eye, and this can produce some curious effects.

One of the most striking is the twinkling of the stars. Then there are unequal effects of refraction due to the trembling and fluttering of the atmosphere caused by winds and air currents. These cause tiny waves in the air (as we shall see, the curse of the telescopic astronomer) which complicate the path of the light, causing the image to dance and quiver erratically. Like a prism, the atmosphere not only bends light from its path but also decomposes it into a tiny spectrum. Thus we see first the red and then the green wavelengths, the violet end of the spectrum being absorbed in the atmosphere. If the agitation of the atmosphere is very great, there may well be a complete extinction of the light from a star for a fraction of a second. In this respect, the atmosphere behaves as if it were made up of millions of trembling crystals, such as are found in cut-glass chandeliers.

This is confirmed by everyday experience. Distant lights such as street lamps can be seen to tremble under suitable conditions. Their light, which comes to us almost horizontally as a rule, usually has to pass through the unequally heated layers of air just above the ground. Distant features of a landscape seen across a plain or a wide road on a hot day will shimmer in the same way. This is connected with the mirage effect, which we shall discuss later.

People are often surprised to learn that the planets do not twinkle in the same way as the stars. The stars, however, are simply pinpoints of light, which dance and twinkle when seen through our atmosphere, whilst the planets have quite an appreciable diameter, and the agitation of the air affects only one part of their image at a time. Some part of the light-pencil will always get through to the observer's eye, and this will produce a steady instead of a twinkling appearance. This, however, is not so in the telescope, which magnifies the 'boiling' of the air as well as the image. Sometimes conditions are so bad that a planet will be no more than a bright blur in the field of view, and then all the observer can do is to close up the instrument and go to bed.

Another striking demonstration of atmospheric refraction may be had by examining the Sun at sunrise and sunset. This should preferably be carried out with the aid of a pair of binoculars or a small telescope, which will reveal details which escape the unaided eye. It cannot, however, be over-emphasised that the greatest possible care should be taken not to look at the Sun through any optical instrument on a bright, cloudless day. Choose

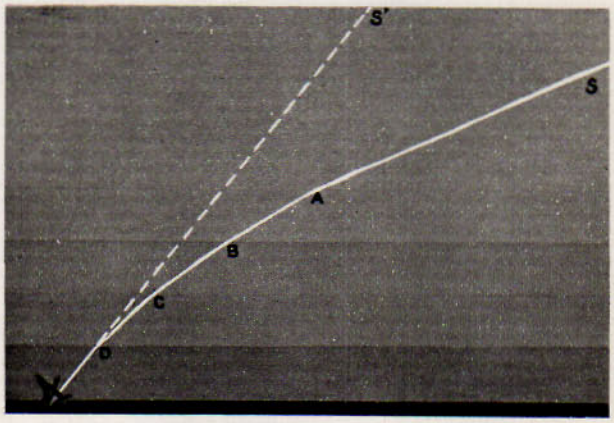


Fig. 216 — PRINCIPLE OF ATMOSPHERIC REFRACTION

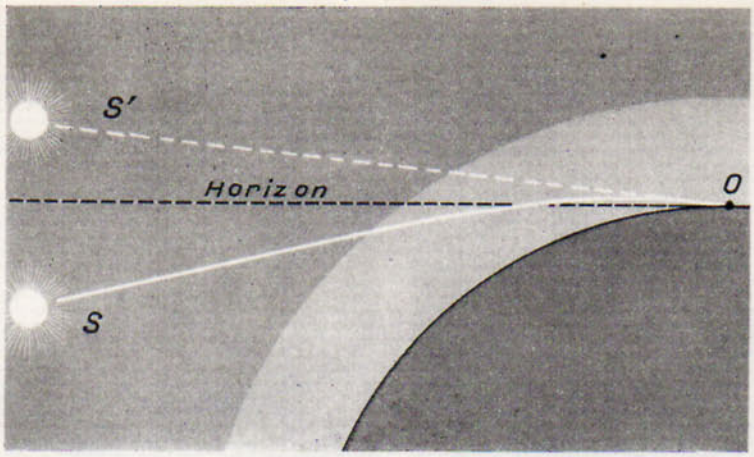


Fig. 217 — ATMOSPHERIC REFRACTION. A star below the horizon may be brought into view by this phenomenon

a calm, misty morning or evening, when most of the Sun's brilliance has been lost in the atmosphere, but even then a dark glass should be used to protect the eye. Better still, set up the instrument on a stand and project the image on to a white card held a foot or so from the eye-piece. This is absolutely safe.

Following what has been said about the increasing value of refraction towards the horizon, it will easily be understood that at sunset, the lower half of the Sun's disc, being nearer the horizon, will be more strongly refracted than the upper. This unequal refraction will deform the disc, flattening it almost into an ellipse, since its horizontal diameter will remain constant (Fig. 220).

It is of course when the Sun reaches the horizon that the effect becomes greatest. When the Sun rises in the morning, the effect diminishes, but it still remains appreciable when the Sun has reached five or six times its own diameter above the horizon.

This general behaviour is further complicated by the various anomalies that come into play. The atmosphere may be full of relatively thin layers of air at different densities (already discussed in connection with the twinkling of the stars). These create surfaces of discontinuity, which may have curious effects. They may distort the circular outline of the Sun in the most capricious manner, as it nears the horizon, whilst waves in the air may detach segments of the solar disc which seem to shatter into fragments of molten gold. Fig. 219 and 221—223 illustrate some of these effects.

The phenomenon of refraction also explains the famous 'green flash' often observed at sunrise or sunset. Acting like a prism, the atmosphere disperses the normal white light into a short vertical spectrum, with the red end at the bottom. Under these conditions, the Sun's disc appears fringed with blue or green along the upper edge and red along the lower, because there is a certain shifting of the images formed by each wavelength of the spectrum. And so when the Sun starts to rise, or is just disappearing at sunset, only a segment of the green fringe shows above the horizon. The green is seen on its own, because the blue and violet are absorbed in the atmosphere.

We shall not dwell on these details, interesting though they are, especially to meteorologists, but it was necessary to point them out as visible evidence of the way in which the atmosphere hinders our attempts to observe the stars. We shall end this section with a discussion of mirages, yet another effect of atmospheric refraction.

We have all heard of the illusions which torment thirsty desert travellers, when they think they can see a beautiful expanse of water across the arid sands, reflecting the distant mountains and the silhouettes of palm-trees, only to find that when they rush towards it, the oasis disappears as if by magic.

Let us remember from the beginning that deserts,

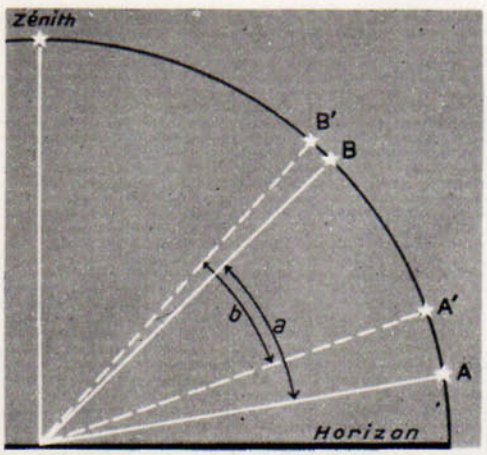


Fig. 218 — THE DECREASE IN REFRACTION WITH INCREASING WEIGHT ABOVE THE HORIZON

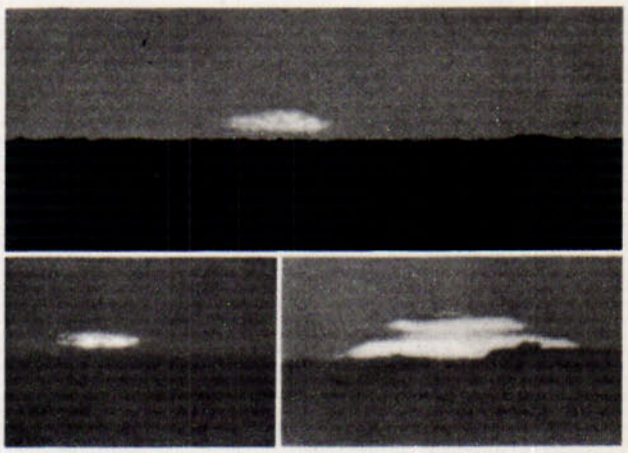


Fig. 219 — LAYERS OF AIR DEFORMING THE LAST SEGMENT OF THE SETTING SUN

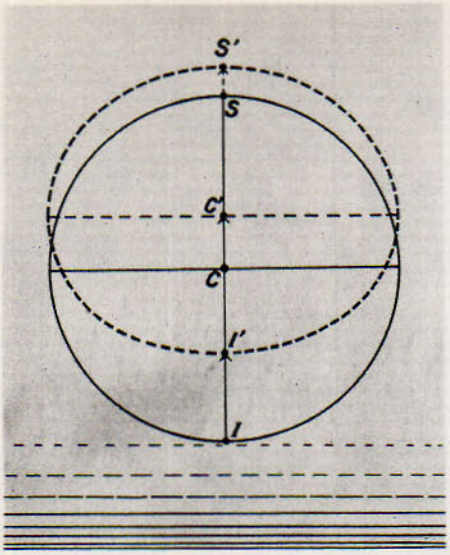


Fig. 220 — THEORETICAL EXPLANATION OF THE SUN'S FLATTENING NEAR THE HORIZON

African or otherwise, have no monopoly in mirages: they are simply more frequent there because the way in which the ground is heated favours them more.

A mirage results from a difference in temperature between the air immediately in contact with the ground and that above. Instead of the less dense layers being at the top, the order is reversed. Thus, a ray of light coming from some moderately high point in the distance such as the top of a hill, a tree or even a building, and entering the decreasingly dense layers of air near the ground, will be refracted along a path which becomes more and more nearly parallel with the surface (Fig. 224). At B it finally comes to a layer where the density is so low that the light is no longer refracted, but totally reflected up along a path which is symmetrically opposite to the one which it has just followed. Thus an observer will see an image of the point A at A′, that is the direction along which the light reaches him, and the reflection of the sky in the layers of hot air will give the impression of a liquid mirror.

As we have said, mirages can be seen anywhere at the right time, both on sea and dry land. They are not confined to any one part of the world, but can sometimes be seen at the poles, since they arise from marked differences in temperature between layers of air on the surface. Thus they are quite common at sea, where a distant range of mountains or a ship on the horizon may be seen, accompanied by its image in reverse, apparently floating on the edge of the waters. This line is, however, a false horizon, the real horizon being hidden by the reflection of the sky. Similar effects can be observed at sunset, when the Sun's disc, which has been deformed by the general refraction, is seen joined to the brilliant streak of its obliquely reflected image. It can be seen on these occasions that the last tip of the Sun disappears not below the true horizon but some distance above it.

Terrestrial magnetism and electricity. — The Earth is the scene of many magnetic and electrical phenomena, of which we may or may not be aware. Around the Earth there exists a constantly varying magnetic field; its variations may be linked with other electrical and magnetic phenomena of the globe, and certain happenings on the Sun.

The existence of terrestrial magnetism is generally known to most people from the magnetic compass, which always indicates the north-south direction. Most of us have also heard of magnetic storms, which betray their presence by changes in the orientation of the compass-needle and by disturbances in long-distance telegraphic and radio communication. To these irregular and sometimes large variations we must add a number of periodic variations: diurnal, annular and secular.

The vertical plane defined by the direction of the compass-needle is the magnetic meridian, which frequently fails to coincide with the geographical meridian of the place. The angle between them is called the declination, either east or west of the north-south line. At the same time, the compass-needle makes a certain angle with the ground: this is called the inclination or dip. These are the factors measured in the study of magnetic variation.

The diurnal variations are small, and appear as oscillations of a few minutes of arc at the most; they themselves vary according to the place of observation, the epoch and the year. The annual variations are much smaller and not so well defined. The secular variations occur slowly and continuously, and are detected by the long-term comparison of records. The most important of these variations is the progressive change in declination; in Britain, for example, where it is east at present, the declination is decreasing by several minutes of arc per year. Since records for London are available for a very long period, we may find it interesting to trace what has happened in past years (Fig. 227).

Fig. 221 — THE SUN'S DISC IS ALREADY NOTICEABLY FLATTENED SOME TIME BEFORE IT REACHES THE HORIZON

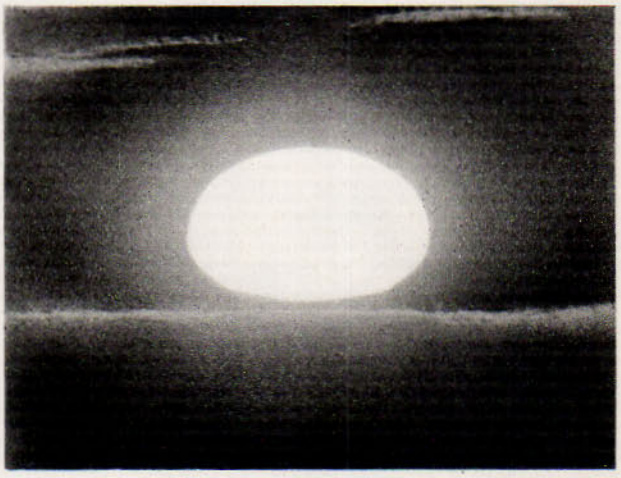


Fig. 222 — FLATTENING OF THE SUN AT THE HORIZON

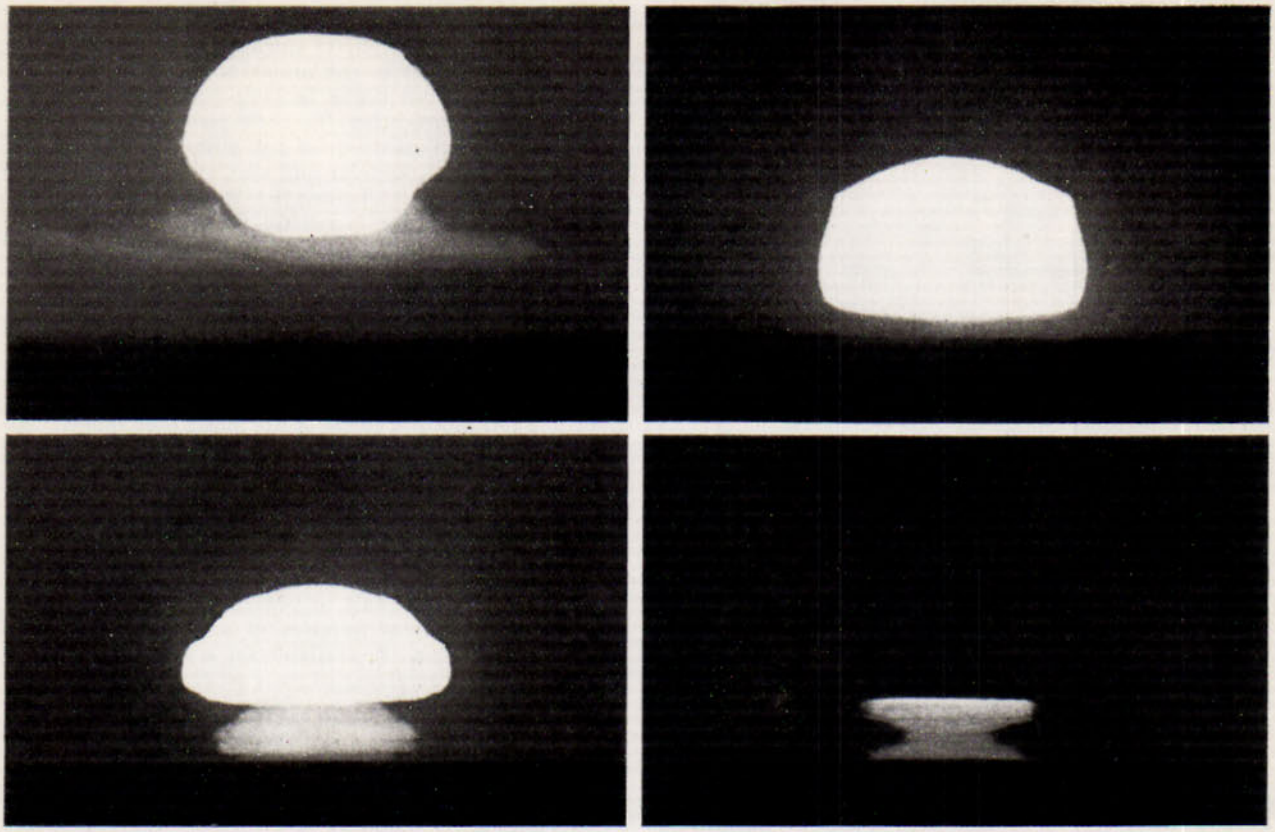


Fig. 223 — CURIOUS ANOMALIES OF REFRACTION OVER THE SEA, DUE TO LAYERING OF THE AIR

The earliest records we have are of observations made in London in 1580, probably coinciding with an easterly maximum, viz. 11°. Subsequent observations showed a steady westward drift; the value passed through zero in 1657 and reached westerly maximum — 24.3° — in about 1815. When the declination will again be zero is a matter of conjecture. General trends, however, suggest that this may be early in the next century, possibly about 2010.

Therefore, in addition, to the variation in the value of declination from place to place, we must allow for the variation according to time in order to be able to use a compass effectively.

If we draw lines on which every point has the same declination (these magnetic contour-lines are called 'isogonics') we trace out a complex network over the globe, extending from one magnetic pole to the other or back round to the same pole. There are some strange anomalies in the arrangement of the isogonics which have not yet been satisfactorily explained, and indeed there are still many unsolved problems connected with the terrestrial globe.

The Earth's magnetic field is arranged as if there were a bar magnet at the centre of the planet, making a slight angle with the axis of rotation. Accordingly, the north magnetic pole would be at latitude 78°5′ and longitude 69° W, that is to say in the north-western part of Greenland, whilst the south magnetic pole would be exactly opposite. The magnetic axis would therefore make an angle of 11°5′ with the geographical axis. However, attempts to locate the exact position of the pole have yielded very different results, especially as the north and south pole are *not* exactly opposite each other. James Ross gave the following position in 1831: latitude 70°5′, longitude 96°46′W. Since then, differences of a few degrees each way have been detected. It is the same for the south magnetic pole: in 1840 its position was given as latitude 70°30′, and longitude 147°30′. Therefore, as we said, these positions are variable, and Chree's discussion of the results obtained in 1912 by two expeditions operating for several months in the vicinity of the south magnetic pole concludes that during this period the pole had wandered along an arc of a five-kilometre circle. Moreover, at times of great magnetic disturbance, the poles may wander from their normal position by anything up to several hundred miles.

These observations correspond with certain solar phenomena, suggesting that the latter may have some degree of direct influence on our world. This very important subject will be discussed at length in Chapter 10.

Magnetic storms are sudden variations in the magnetic field; sometimes the disturbances are considerable. These remarkable storms often take place on a very large scale, even embracing the whole Earth on occasion. They are frequently accompanied by beautiful displays of the Polar Lights, which we touched upon earlier.

We may first note that the name 'Northern Lights' (or Aurora Borealis) refers to the phenomenon as it appears in our hemisphere: in the other it is called the Aurora Australis. Because the former are the more frequently seen since there are greater numbers of people living in high northern latitudes, there is a tendency for the use of the term to become more general, and applied to aurorae anywhere. Moreover, although these aurorae frequently appear in high latitudes, where they can be seen to full advantage, they do not belong exclusively to the polar regions, as their name would imply, but spread on occasion over enormous areas of the Earth, sometimes right down to the equator. The appearance of an aurora in Britain is in no way exceptional: it is just sufficiently

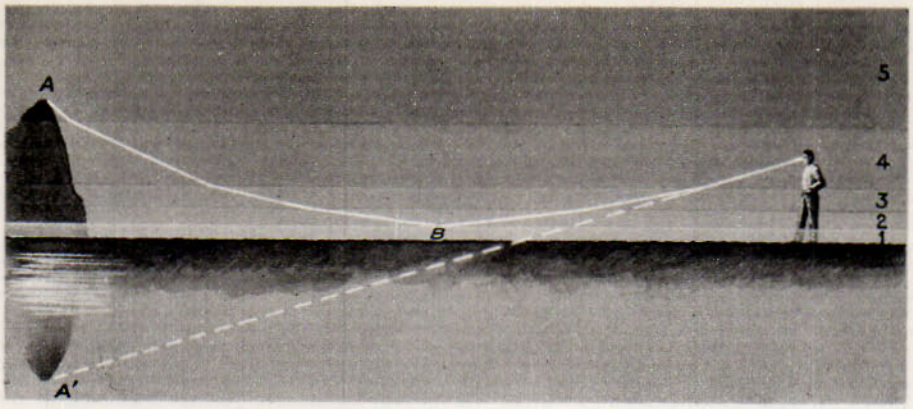


Fig. 224 — EXPLANATION OF MIRAGES

Fig. 225 — PHOTOGRAPH OF A MIRAGE

rare for people not to be familiar with the phenomenon, but a careful observer can record quite a large number during the course of a year. Usually they take the form of a faint glow suffusing the sky just above the northern horizon, sometimes they are moving patches of light, or vertical rays like searchlight beams, occasionally disappearing momentarily, and often flickering. Where they are most plentiful, in the polar regions, they light up the sky with a splendid brilliance which drives away the winter darkness, often taking on weird and beautiful shapes: twisting flames, waving curtains of cold fire. No words can do justice to their beauty for someone who has never seen them.

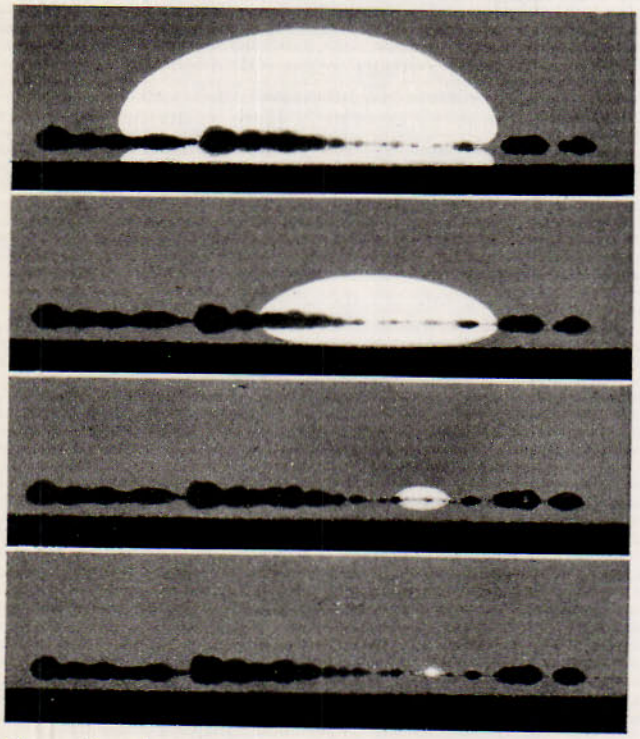


Fig. 226 — A MIRAGE MAY ALTER THE APPARENT POSITION OF THE HORIZON

The explanation of this phenomenon, which seemed so mysterious for such a very long time, can be found in the streams of charged particles, in this case electrons, coming from the Sun. They were first recognised by Deslandres and Birkeland. P.Villard devised an auroral theory derived from his many experiments with vacuum tubes; according to him, the phenomenon was due to rays moving under the effect of the Earth's magnetic field. Their trajectories, which converge on the polar regions, were studied by Störmer and found to be in accord with the many spectacular observations of the phenomenon made by the famous Norwegian scientist. Moreover, his calculations apply to all other sorts of charged particles, whether negative or positive. Whatever the exact nature of these particles, their arrival in the upper atmosphere causes a luminous display through the bombardment of molecules in this ultra-rarefied medium. The phenomenon is similar in appearance to what happens in a Crookes tube, when an electric discharge through a rarefied gas causes a glow.

Apart from the many phenomena which will be discussed in the chapter on the influence of the Sun on the Earth, there are others which affect the electrical properties of the atmosphere. These properties concern the air's electrical conductivity and the currents to be found in it. The study of these extremely complex problems reveals features which must be dealt with more specifically. In this present account of the conditions existing on our world and its physical activity, we are bound to keep to

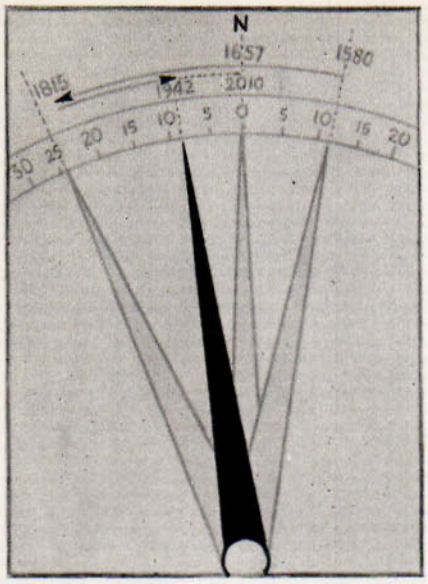


Fig. 227 — SECULAR VARIATIONS IN THE MAGNETIC DECLINATION AT LONDON BETWEEN 1580 AND 2010

the principal facts and more particularly to those which are directly in evidence. Yet these explanations require us to be familiar with certain fundamental ideas.

First of all, the Earth is a negatively charged electrical conductor. The intensity of the electrical field decreases with distance all around; from 100 volts per metre at ground level, it falls off to only a few volts per metre at a height of 4 to 5 miles. At very great altitudes, however, these potential values rise steeply.

Areas of the same potential are distributed in bands approximately parallel with the surface of the globe; but in actual fact, they can be distorted into many complex shapes by the relief of the ground or by buildings. Thus above a valley between two hills, these bands may be wider than they are above the hills on either side, where the electric field is more intense. A similar phenomenon is found to occur round tall conducting buildings such as the Eiffel Tower, at the top of which the potential gradient is a hundred times steeper than at the ground.

The difference in potential between the ground and a point in the atmosphere is usually positive, and increases with the height of the point. The intensity of the field varies continuously, undergoing periodic variations, both diurnal and annual. There are also very large random variations. Thus in temperate regions, the field which, on average, is about 100 volts per metre, varies from 75 in summer to 135 in winter. Intensity is on the whole slightly lower at the equator, but at the poles is not very much higher than in temperate latitudes.

The vicissitudes of climate also have their effect on the variations in the electric field, but we do not yet understand what this signifies. In particular, the wind transports ions in the atmosphere, and also surface particles such as smoke and various types of mineral dust which change the conditions of ionisation. In the same way, the water-droplets of mist and rain carry electric charges, which cause certain perturbations; those associated with snow are even stronger, whilst storms cause intense disturbances.

We are much more impressed by the spectacular nature of storms than by the important but imperceptible forces continuously present all around us. However, storms bring us to consider the subject of atmospheric electricity. These violent phenomena are connected with those immense cumulo-nimbus clouds containing very strong updraughts. With their bases about 3,000 or 6,000 feet off the ground, they often rise to a height of 15,000 or 16,000 feet, and they may be thought of as tremendous electrostatic generators, which are working even when most of us would think they are not, because we can see no lightning discharges, the sure sign of a storm.

We may wonder if they work in the same way as a machine which generates static electricity through friction. Whether or not this is true, it is known that electric charges detach themselves from the heart of the cloud and move towards the summit or the base, which become highly charged with negative or positive electricity. This causes an intense electric field within the cloud, between the cloud and its neighbours and between the cloud and the ground. Under these conditions the negative or positive strength of the electric field in the vicinity of the ground may become extremely great, perhaps several thousand volts per metre, and high points may begin to give off a luminous discharge.

When these electric fields reach sufficient intensity, there occur the discharges which we know as lightning. They strike within the cloud, from cloud to cloud or from ground to cloud. The length of a lightning flash may be considerable: from the cloud to the ground it is only a few thousand feet, but between clouds it may be anything up to ten miles. Their frequency is sometimes so great that in some storms discharges follow at intervals of only a few seconds, while in exceptional cases they may occur at a rate of more than one a second.

Seen with the naked eye, lightning flashes usually look like a zigzag, and this is the way in which they are nearly always drawn, sometimes in the hands of Jupiter, sometimes as the emblem of the electrician.

It was only with the use of photography that this erroneous impression was corrected. Forked lightning in fact follows a sinuous path broken by no sharp angles, and usually consists of a main streak, rather irregular in shape, with a number of ramifications. Modern investigation has shown that whereas a flash of lightning is a single blinding streak to the naked eye, it is actually formed of an ultra-rapid succession of five or six streaks following almost the same path. Here we are dealing with the principal streak, whilst the ramifications occur only during the first two or three discharges. It is probable that the intense ionisation of the air at the first discharge determines the path of the others.

The duration of these discharges is no doubt variable, but must certainly be extremely short. Various estimates give it as about one ten-thousandth of a second, or one thousandth at the most. This means that the succession of flashes we see as a single one lasts from about one thousandth to one hundredth of a second. Visually, the multiple nature of a 'lightning flash' cannot be detected because the bright image persists on the retina. The zigzag appearance is probably due to the fact that a flash of lightning occurs as a series of flickers.

Lightning flashes literally blind one's eye at night. They are so bright that for a tiny fraction of a second the landscape looks as light as day. This is what happens in the case of 'summer lightning'. These noiseless, diffuse flashes are in no way different from the usual manifestations of electricity we hear thundering overhead during storms. Thunder is simply the noise produced by a lightning discharge, a hard crack when the storm is overhead and a low, rumbling reverberation when it is at a distance. If the storm is too far away, the thunder will die out before reaching us, whilst the intense lightning flashes will still be seen. 'Summer lightning' is in fact merely the

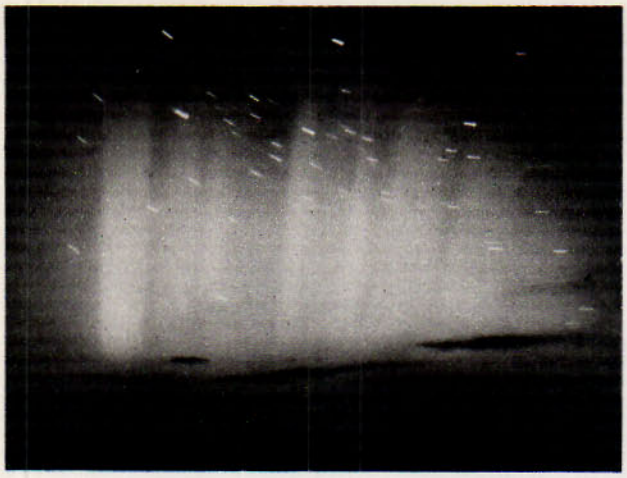


Fig. 228 — AN AURORA BOREALIS (RAY TYPE)

Fig. 229 — AN AURORAL GLOW (LOOKING NORTH)

reflected light from a distant storm, the actual discharge flash being hidden by clouds or the horizon.

The only other type of lightning is that which is known as spherical lightning, the discharge taking a globular form. We know very little about this phenomenon, and can explain it even less.

In all these electrical activities, the amount of energy involved is extremely great. This we can tell merely by looking at the damage caused to anything struck by lightning. Measurements indicate the presence of currents of about 10,000 amperes or more.

All this information is derived from our studies of the lower regions of the atmosphere, and about the higher reaches we have very little knowledge. We do know, however, that the ionised layers mentioned earlier are very good conductors, and it is quite possible that they are the seat of strong electric currents which may have considerable effect on the Earth's magnetic field.

Back on the surface we find more electrical activity, this time known by the name 'telluric currents'. These phenomena, not well understood but certainly very complex, can be detected by burying two metal plates in the ground at some distance apart and joining them with a wire. A galvanometer connected to the wire reveals the existence of electric currents, constantly varying in direction and intensity.

There seems to be some relation between the variations in these currents and those in the Earth's magnetic field, and the telluric currents become particularly strong during intense magnetic storms; but their exact nature is still in doubt. Some workers incline to the view that they are induced currents due to variations in the magnetic field, but this seems to be invalidated by observations which show that variations in the telluric currents generally precede those in the magnetic field; according to J. Bosler, the magnetic disturbances are caused by telluric currents.

To sum up, we can say that all this electrical activity, both inside and outside the globe, is very complex and varied; all the different sorts of activity seem to be connected, in ways which we do not understand. There are indeed many problems still to be solved, and they will take years of patient and interesting research. For example, the explanation of terrestrial electricity founded on ionisation is a very recent idea, and many points remain to be cleared up. But we already know enough to give a satisfactory interpretation of the conductivity of the air and of the origin of the electric charges observed in the air. As we saw earlier, the Sun's influence is of prime importance, but it is evident that the radioactivity of the ground also plays an important part in ionisation, at least where the lower layers of the atmosphere are concerned, whilst the upper atmosphere is more affected by influences from outside the Earth.

The majority of rocks in the Earth's crust contain radioactive elements such as uranium and thorium; their disintegration products are liberated into the atmosphere, but in unequal quantities at different places on the Earth's surface. Above the land masses, this emission is enormously greater than over the seas, since the radioactive material content of the oceans is extremely small, being no more than a few thousandths that of the continents. Therefore any radioactivity in the air over the sea must have been borne there on air currents from the continental areas.

The water of lakes and rivers is hardly richer in radioactivity than the seas, but springs, especially hot springs in volcanic regions, show a very marked activity, and the ionisation of the atmosphere is higher in their neighbourhood.

There are other kinds of activity going on around the Earth, but it is difficult to evaluate their importance. There is the activity of waterfalls and the turbulence of the waves, the production of ions during combustion, the emission of electric charges by solid particles suspended in the air and bombarded by solar radiation (especially by ice-needles in cirrus cloud).

Then there are the principal facts and phenomena revealed by the electro-magnetic properties of the Earth. We still have to explain the origin of the globe's negative charge which must be postulated in our fundamental discussions, and we do not yet understand how it is maintained. Many hypotheses have been suggested, including the existence of a highly penetrating stream of negative charges coming in from space, which pass through the atmosphere with no appreciable absorption and impart their charge to the ground.

Many of those matters would possibly be clarified if we knew a little more about the interior of the Earth, which is the next subject for discussion.

The Earth's crust. — It is an obvious fact that, whereas we can explore upwards through the Earth's atmosphere directly, by means of sounding balloons, rockets and now Earth satellites, for studying the structure of the planet itself we are forced to rely largely on indirect, derived information. Thus our picture of the Earth's interior changes radically as new techniques lead

Fig. 230 — A CUMULO-NIMBUS CLOUD IS A POWERFUL ELECTROSTATIC GENERATOR

to the formulation of new hypotheses. Let us therefore first survey the situation as it was prior to the most recent developments.

It is generally agreed that for a certain depth the hard rocky layers form a sort of skin, called the Earth's crust. Lower down, the physical properties of the rocks change, and they become more plastic, so that under prolonged stresses deformation can occur. The substance composing these lower layers, about whose constitution we are still uncertain (although the most recent theories offer a satisfactory idea), is known as 'internal magma'.

It was formerly thought that the crust uniformly covered the interior of the Earth, which was supposed to be molten. Support for this theory was found in the gradual rise in temperature at increasing depth, which, if maintained, would mean that beyond a certain depth no rock could remain in the solid state. The region where the 'fluid' state was supposed to begin lay between 50 and 100 miles beneath the surface. At this thickness the crust — compared with the 8,000-mile diameter of the Earth — would be like an extremely thin skin, and it was often compared to an eggshell.

We shall return to this question of the rise in temperature with depth in a moment. However, this concept of the fluidity of the Earth's core was open to a number of objections, as was the idea of its extreme temperature. It was of course largely figurative, for the pressure increases steadily towards the centre of the globe, and at the 'fluid limit' would already be so enormous as to deprive the word 'fluid' of most its ordinary meaning. It is impossible to imagine the state of a body at a temperature of several thousand degrees and under pressure sufficient to make it solid.

In any case, the idea of a fluid interior proved irreconcilable with the well-established explanation of the tides by the attraction of the Sun and Moon, which we shall study in Chapter 6. A fluid interior ought to move in the same way as the liquid ocean covering it, and this tidal motion ought to be detectable in the overlying crust. However, the fact that we have not observed any such effects must not be taken as indicating that they do not occur at all, and that the Earth is completely rigid. On the contrary, both the astronomical and geophysical evidence suggest that the Earth behaves as an elastic body, constantly being deformed by internal and external influences. It is only that the degree of elasticity is very small.

In 1878, after analysing a great deal of evidence about

Fig. 231 — LIGHTNING STRIKING BETWEEN CLOUDS

Fig. 232 — LIGHTNING STRIKING TO GROUND

the tides, Lord Kelvin concluded that as a whole the globe has an elasticity comparable with that of steel. By and large this estimate is supported by the variation of latitude due to the wandering of the geographical pole by a few yards on the surface; by experimental studies of daily variations in the vertical; and by the observations of Michelson and Gale, carried out in 1919, of the fluctuation of the level of water contained in very long tubes buried in the ground. All these show that slight tidal effects occur in the crust, with an amplitude of about 13 inches on the average. If the globe were fluid throughout, the amplitude would be 20 to 24 inches.

But in any case the whole concept of the Earth's 'central fire' must be discarded, as there are insuperable objections. For reasons detailed below (*The Internal Temperature of the Earth*), it is now known that the interior of the Earth cannot be much hotter than the surfac layers.

Vulcanism. — All this is hard to reconcile with the very natural idea that volcanic activity draws its energy from an immense incandescent core inside the crust. Volcanic eruptions were first thought to be caused by the expansion of molten lava after the infiltration of water, which would immediately be turned into large quantities of steam. The tremendous pressure of this steam was then thought to burst through the crust (especially at weak points such as faults and folds), liberating large quantities of fluid magma, water vapour, gases and debris of various sorts. This logical theory was apparently supported by the distribution of volcanic activity, which is principally found near lines of mountain folding along the great ocean depressions.

New work, however, has revealed the existence of different conditions in the interior. Different volcanic lavas, even when found comparatively close together, do not have the same composition. It would seem therefore that volcanic activity results from great pockets or reservoirs of lava in the crust, in which violent and as yet unknown chemical reactions take place. Radioactivity is now held to be the source of heat. In all these cases, however, we are still left to explain the starting and stopping of eruptions.

In this connection, certain workers have tried to establish correlations between solar activity and various disturbances in the physical nature of the Earth, but little success has been reported.

Volcanic eruptions give off certain materials: scoriae, lava, ash and rocks. Scoriae come from molten lava which, when blown to fragments by the pressure of contained gases, cools down to form small lumps of cindery material. Ash comes from the pulverisation and subsequent rapid cooling of lava. It is composed of extremely fine vitreous particles and small crystals (sometimes broken) of iron oxides, pyroxene, etc. Being so fine, it may remain suspended in the air for long periods, travel vast distances and rise to great heights. After the eruption of Krakatoa in 1883 (the greatest cataclysm of this type ever recorded) the atmosphere was 'dirty' for several months, and traces of the dust remained even after years. This, however, is not a unique case; the eruptions of many volcanoes provide similar examples on a smaller scale.

Further evidence of the violence of volcanic eruptions is furnished by the projection of boulders weighing several tons to heights of several thousand feet. The most important volcanic effects, however, are the sheets of lava left behind. Their depth and the area they cover obviously depend on the conditions under which the eruption occurs, the viscosity of the lava and the slope of the

Fig. 233 — 'SUMMER LIGHTNING'

Fig. 234 — THE VOLCANO ASAMA (JAPAN) IN ERUPTION *(NYT photo)*

ground. The last two will determine how rapidly the lava flows, as it cools down from its initial temperature of 2,000 or 3,000°C. The speed of a stream of lava is illustrated by one which poured out of Etna, flowing at the rate of about a mile in three or four days, destroying and submerging everything in its path. As for the area covered, Mauna Loa (Hawaii) has emitted lava-flows more than 30 miles long by 200 yards wide and 100 yards thick at some points.

Earthquakes and the interior of the globe. — Seismic phenomena — always impressive and often tragic — are given the general name 'earthquake' although they may be due to different causes. Some are the result of volcanic activity which is violent enough to cause shocks in the ground; others may be the result of earth subsidence and collapse caused by the action of water. These, however, are localised round the point of occurrence. Earthquakes proper, especially those of catastrophic magnitude, are known to be due to tectonic causes, the mechanism of which may be summarised as follows.

Somewhere within the crust, tensions are set up; these cause strains which are finally relieved by some rupture of the crust; the spot where this occurs is called the focus. Around this point, geographical effects occur almost immediately, perhaps over hundreds of square miles of territory. These are the major shocks, but they are accompanied by minor shocks which would pass unnoticed but for being picked up by specially designed instruments called seismographs, which record their magnitude, direction and character. As some indication both of the sensitivity of these instruments and of the number of earthquakes, we can say that whilst about five thousand shocks can be felt yearly, of which only about one hundred are destructive, our instruments reveal a total of about ten thousand per annum. In certain areas, such as the Japanese islands, the earth trembles constantly. The information derived from the study of earthquakes is an invaluable indication of the interior condition of the Earth.

Fig. 235 — THE AFTERMATH OF AN EARTHQUAKE: KITAMA, JAPAN
(Photo, J. Boyer)

Fig. 236 — A SEISMOGRAM, THE 'SIGNATURE' OF AN EARTHQUAKE

This information is gained through studying the behaviour of shock waves travelling from the focus through the Earth. There are three sorts of waves:

Longitudinal, which deform the medium through which they are travelling in the direction of travel;

Transverse, perpendicular to this direction;

Surface, resembling the waves caused by throwing a stone into a pond.

These waves arrive at the observing station in a certain order according to their nature, and are designated by the letters P (primary), S (secondary) and L (long).

Before dying out, the long waves may accomplish a complete journey round the Earth at an almost constant speed of 3,500 metres per second. From L waves we have learnt that the focus is fairly near to the surface: within 15 miles or so. The P and S waves behave differently; they plunge down into the depths of the Earth with mean speeds of about 7,000 and 4,000 metres per second respectively in the early stages, but as they go deeper these speeds increase according to the laws governing velocity in a medium of constantly increasing elasticity. These waves are propagated from one side of the Earth to the other along various paths: some almost direct, some refracted, some reflected (Figs. 237 and 238) by contact with layers of different density.

From the distance between the epicentre (the nearest point on the surface to the focus) and the observing station, the character of the waves and their time of travel, we can deduce the course they must have taken inside the Earth, which is the same as finding out to what depth they have penetrated because of refraction, etc.

The importance of the light thrown on the nature of the Earth's internal structure by this technique cannot be over-emphasised, and the results obtained are largely responsible for our current picture of the constitution of the Earth. But before turning to this aspect of the subject, let us deal with the question of the planet's internal temperature, which remained a very vexed one for a long while, but now seems to be receiving a satisfactory answer.

The internal temperature of the Earth. — The ground is warmed by radiation from the Sun, but this varies according to the time of day and the season, the latitude, geographical position and topography of the region. Variations in the amount of heat received occur mostly at the surface and progressively less further underground. Below about 3 feet from the surface, the effect of rapid diurnal changes is not felt at all, and seasonal change has no effect deeper than about 70—100 feet, according to the nature of the terrain. A constant temperature might be expected to reign beneath this limit. But, as we have said, the temperature actually rises with increasing depth, and there must be some source of heat-energy to account for this.

The depth required for the temperature of the ground to rise one degree C. is called the 'geothermic degree'. Its value is not constant; even over small areas, quite considerable variations can be measured. Thus in the Anzin mines the geothermic degree varies between about 50 and 84 feet in different shafts, and these values are themselves lower than average. The record seems to be 23 feet 8 inches at Oberstaten (in the oil-bearing region of Alsace) but elsewhere it may be as much as 210 feet. Attempts have been made to establish a 'mean geothermic degree', taking the average of a large number of measurements over a very wide area, and these seemed to give a result of about 100 feet for Europe, and 136 feet for the USA. Taking the world as a whole, a mean value of about 107 feet was indicated.

If this value were constant, at a depth of 60 miles the temperature would be 3,000°C. As mentioned earlier, the most refractory substance would be in a molten state at this depth, at least in the laboratory sense of the term.

But the wide variation in the value of the geothermic degree, even over very short distances, led to suspicion being cast on this explanation of its cause, and the old idea of a 'central fire', when closely scrutinised, turned out to be quite untenable. The different estimates which had been made of the Earth's internal temperature varied widely since each one of them was largely dependent on the nature of the ground where investigations were carried out. They all resulted in an attribution of very great heat to the core, however. Under the very high pressures which must obtain at the planet's centre (of the order of 4 million atmospheres) the thermal conductivity, which increases rapidly with rising pressure, will likewise

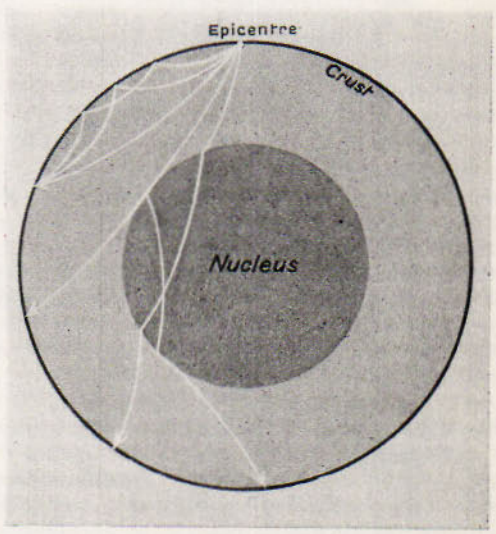


Fig. 237 — PROPAGATION OF SHOCK WAVES INSIDE THE EARTH

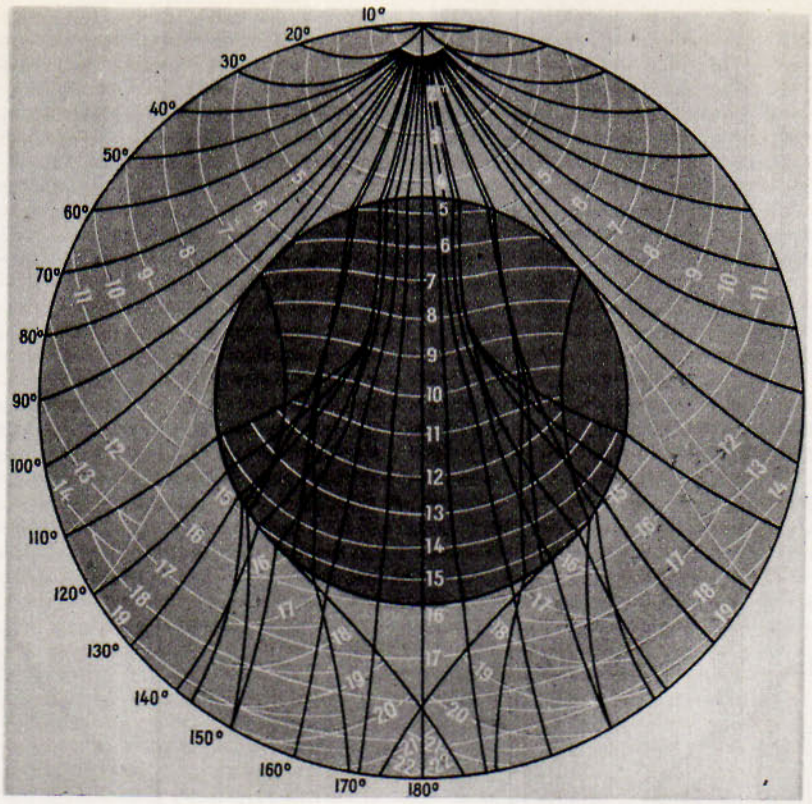


Fig. 238 — THE PROPAGATION OF DEEP-SEATED SEISMIC WAVES, ACCORDING TO GUTENBERG. The diagram shows the path of the compression waves, some of which are reflected and others refracted by the core. The time taken (in minutes) may by read off the scale of longitudes. It will be seen that the waves reflected by the core arrive at a point 120° from the earthquake in 15 minutes, whilst those refracted by it take 19 minutes. At 150° from the earthquake, the waves refracted by the core and the central nucleus arrive within 1 minute of each other.

attain a very high value. Thus if any heat is produced in the Earth's interior it will rapidly escape. So it is certain that the temperature difference between the Earth's centre and its surface layers is not large.

The origin of the Earth's heat cannot be explained in terms of incomplete cooling from an original molten state. Such cooling, when it occurred, would not have occupied more than a few million years. Now the age of the Earth is today well established — it is in the neighbourhood of 3.5×10^9 years, long enough for the cooling to have taken place many times over.

It is almost universally agreed nowadays that the greater part of the Earth's internal heat is produced within the crust itself by the natural radioactivity of rocks containing (sometimes in very minute quantities) radioactive minerals of the actinium, radium and thorium families. All the phenomena of vulcanism and internal heat are attributed to this agency.

Having disposed of the old idea of a thin solid crust floating on an incandescent core, we shall now have to explain the picture of the Earth's structure which has replaced it and the reasons for adopting this new hypothesis.

The composition of the Earth. — We have already discussed the mass of the Earth, determined by experiments in which we measure the gravitation constant. The value thus arrived at for the mean density is 5.52, water being taken as unity.

The values derived from the analysis of many different sorts of rock found in the crust are much lower, being about 2.28 on average. It therefore follows that in the central region of the Earth the density must be much higher than 5.5, hence comparable with the metals. If the density rose constantly from the surface inwards, it would be 11.6 at the centre. However, seismological studies (a tool for the investigation of the Earth's interior the importance of which it is impossible to over-stress) have indicated that the density is not distributed uniformly by any means.

The principal facts needing to be taken into account when constructing a picture of the Earth's interior were established by seismology. Very briefly summarised, they are as follows:

(1) A marked discontinuity in the velocity with which P waves are propagated occurs at a depth of 1,800 miles.

(2) Beyond this same level, S waves are propagated no further; this is a feature which essentially characterises liquids, and it must therefore be concluded that the Earth's core behaves as though it were liquid, at any rate so far as the general characteristics of its atomic distribution and separation are concerned.

(3) A second discontinuity has been found to occur at a depth of 3,000 miles, and still another (much less marked) at about 60 miles.

(4) These discontinuities show themselves by the reflection of the P and S waves and the partial refraction of certain of the P waves. The effects of these reflections can be detected, though not without formidable difficulties, from the analysis of seismograms.

Knowing (a) the depths at which these discontinuities in propagation velocities occur, (b) some of the velocities

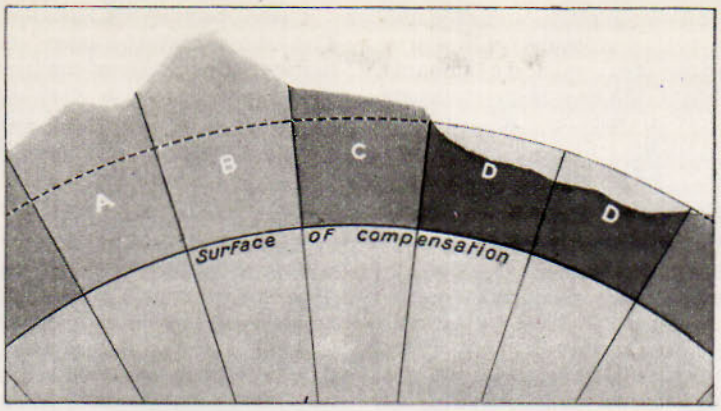


Fig. 239 — THE PRINCIPLE OF ISOSTASY

themselves, (c) the mean density of the surface rocks (about 2.7), (d) the density of the intermediate underlying layer (3.3) and the mean density of the Earth as a whole (5.5), we have sufficient data to investigate the probable distribution of densities throughout the Earth's interior. One other factor that has of course to be taken into account is the coefficient of compressibility of the rocks, which can be determined in the laboratory.

In this way, different 'models' have been proposed, and it is a remarkable fact that all those that have been discussed by the leading geophysicists resemble one another in their essentials.

First comes a granitic surface layer, some 20 miles thick; the first discontinuity occurs below this at a depth of 60 miles (256 miles according to some authorities); the density then increases to about 5 at a depth of some 1,800 miles, where it rises sharply to 9.5, thereafter increasing still further, to 11.5, between 1,800 and 3,000 miles; at 3,000 miles it again rises sharply, to 17. This second, inner core transmits S waves, behaving in this respect like a solid.

Satisfactory agreement seems likewise to have been reached regarding the constitution — not merely the distribution of densities and pressures — of the Earth's interior. We have seen that modern theories have revised the idea of a 'central fire'; another hypothesis that has had to be abandoned — though it shows itself equally tenacious of life — is that of a metallic nucleus. According to certain geologists this nucleus consists of almost pure nickel and iron, and contributes about 32% of the Earth's total mass.

This hypothesis, which was proposed by Wiechert in 1897, rests on the following two considerations:

(1) The density of iron (7.8) might very well be increased to 10 under the influence of high central pressures.

(2) A certain proportion of meteorites — the so-called 'iron' meteorites — consist of an almost pure nickel-iron alloy. (The other main type of meteorite is that known as 'stony' — see p. 256.)

The division of the Earth's body into layers of this sort is quite inconceivable. It would require several 10^{12} years for such divisions to become even as well established as that whose existence at a depth of 1,800 miles is suggested by the seismological observations of the discontinuity at this level. Attempts have therefore been made to interpret this discontinuity, not as a change of composition, but rather as a change of state (brought about by pressure) in the more or less homogeneous material of which the Earth must be constructed.

It is perfectly logical to assume that the original composition of the Earth was similar to that of the Sun, which is in fact well established (see Chapter 10) and known to consist mainly of the lightest elements (hydrogen and helium). This primeval composition would have been profoundly modified, however, by the more or less complete escape of the light gases at a later date. The problem has been thoroughly examined by astronomers, taking into account the kinetic theory of gases (the velocity of the thermal agitation of the molecules, which is a function both of temperature and of the molecular weights of the gases concerned) and the values of the parabolic escape velocity of each planet. The latter quantity is proportional to a planet's mass and inversely proportional to its radius; in the case of the Earth its value is 37,000 ft/sec (approximately 7 m.p.s. or 25,000 m.p.h.). A simple calculation will show that the Earth would have been unable to retain any appreciable quantity of the lightest elements — particularly in the early days of its history, when the temperature was higher than it is now. The Moon and Mercury, on account of their small masses, are incapable of retaining any atmosphere at all; the Martian atmosphere, for the same reason, must be very much attenuated. The four giant planets, on the other hand, will have retained the whole of their original supply of the light gases.

Now it has been demonstrated theoretically as well as experimentally that any body subjected to a high enough pressure will pass into a 'metallic' state, which is characterised by improved thermal and electrical conductivity. This results from the ability of the electrons passing among the atoms to transmit heat and electricity, and it is this free movement of electrons that confers on the body some of the characteristics of a metal.

The application to the Earth of this theory of the passage into a 'metallic' state is known as Ramsey's hypothesis. This neat and satisfying hypothesis explains the average composition of the present-day Earth as having been derived from that of an original nebula through the loss of virtually all its light elements, and furthermore accounts perfectly for the sharp discontinuities revealed by seismology.

Bearing in mind the fact that the critical pressure at which the change of state occurs is different for different substances, Ramsey believes that the observed phenomena can be explained by assuming that one only of the more abundant materials composing the Earth has undergone this transformation — silica, which is an important constituent of the outermost layers of the crust. This consideration lends weight to the suggestion that the material composing the globe is largely uniform.

The actual analysis of the surface layer of the crust yields these data: oxygen, 47.3%; silicon, 27.7%; aluminium, 7.8%; iron, 4.5%; calcium, 3.4%; magnesium,

2.2%; potassium, 2.4%; sodium, 2.4%; titanium, 0.4%; hydrogen, 0.2%; carbon, 0.2%; chlorine, phosphorus, sulphur, barium, manganese, etc., in trace amounts. Since in this layer silicates of aluminium predominate, it has been given the name of sial (*Si+Al*), and has a density of 2.8 to 3.0. Beneath the silicates of aluminium, and having a density of 3.0 to 3.4, is a layer of silicates of magnesium, or sima.

Gravity and study of the Earth's interior. — The study of gravity also contributes to our research into the composition and structure of the Earth's internal layers. At the beginning of this chapter, there was a discussion as to the mass of the Earth and its effects, as determined by the size, shape and behaviour of the globe.

The variation in gravity with latitude may be calculated with the appropriate formulae. When the observed values at any particular place are compared with the theoretical ones (that is to say, reduced to the geoid), there are often considerable differences, known as gravitational anomalies. Apart from certain purely local anomalies that we can allow for, they consist of a general decrease in the force of gravity in mountainous areas and an increase over the oceans. These differences are explained by a relation between differences in the density of the crust and particular features placed on top of it. In other words, the Earth's crust contains some sizeable irregularities which make it thicker and less dense beneath the continental masses than beneath the ocean depressions.

These observations have led to the theory of isostasy. This states that the mass contained in similar cones starting at the centre of the Earth is always the same, and that at a certain depth there is a theoretical surface, concentric with the terrestrial ellipsoid, which everywhere supports the same pressure of overlying masses. In a word, there is the same weight of matter in all conical section having their lower base on this surface (called the surface of compensation) and their upper base at the surface of the globe at places of varying heights. This means, of course, that their volumes will be different (Fig. 239).

Beneath the continents, therefore, matter would have to be less dense in order to fill the greater volume, especially when the terrain is mountainous. On the other hand, where the thickness is less, i. e. beneath the oceans, compression would cause an increase in mass.

This conception, which is of interest both to geologists and geophysicists, has encouraged many workers to try to determine the depth of the theoretical surface of compensation; the Geodesic Service of the U. S. A. arrived at the figure of 70 to 75 miles, whilst W. Bowie made it 60 miles beneath the surface.

Thus the Earth is not simply that enormous mass, immutable and solidly fixed at the centre of the universe, which it seemed in the childhood of our race. Moving through space, subjected to all sorts of influences, our world has a very active physical life, to which the phenomena we have just been studying give us some clue.

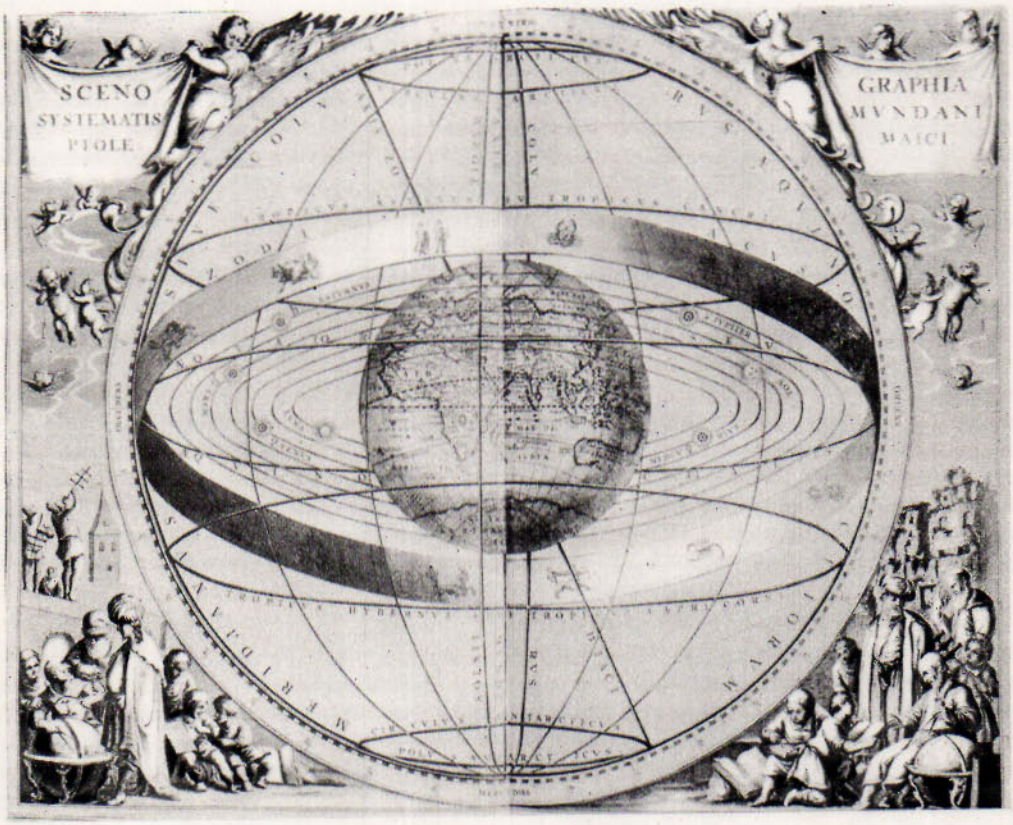


Fig. 240 — THE ANCIENTS IMAGINED THAT THE EARTH WAS THE CENTRE OF UNIVERSE. Engraving from *Harmonica Macrocosmica*, by A. Cellarius (Amsterdam, 1660), showing the Ptolemaic system (second century A. D.). *(Larousse photo)*

Fig. 241 — THE MOON, SEEN LIKE THIS IN THE TWILIGHT, GIVES THE CLEAR IMPRESSION THAT IT IS A GLOBE LIT BY THE SUN

CHAPTER 5

THE MOON

As the only natural satellite of the Earth, the Moon is the nearest of the heavenly bodies; all the rest are extremely distant compared with it. This relative proximity, combined with its situation with regard to our world, has long made it an object of special attention for human observers. The size of its disc, the regularity of its phases, which serve as a measure of time, its soft, poetic light dispelling the blackness of night, have all made the Moon universally familiar and loved.

Moreover, the Moon is the only body near enough for its surface details to be observed through the telescope distinctly and minutely, and nothing is more fascinating than a voyage of exploration on this world which, although lying 'on our doorstep', is so different from our own.

Size and distance of the Moon. — The fact that the Moon is so well seen with the unaided eye is due to its nearness and size. Its diameter is 2,160 miles, which is just over a quarter that of the Earth; more exactly, taking the Earth's equatorial diameter as 1, the Moon's is 0.272.

Its surface area is only 1/14 that of the Earth, whilst their volumes are in the ratio 1:50. The inequality of the two bodies becomes more pronounced when we compare their masses; being composed of materials of lower density, the Moon's mass is only 1/81 that of the Earth. Although the Moon may seem very small in comparison with the Earth, it must be pointed out that compared with the normal relationship between satellites and their primaries, the Moon is very large. Indeed, relatively speaking, it is the largest of all satellites.

In the sky, the Moon has an apparent angular diameter of 31′ 7″, but this is only a mean value. As we saw earlier, astronomical distances, whether between a planet and the Sun or a satellite and its primary, are given as mean distances because all such bodies move in elliptical orbits,

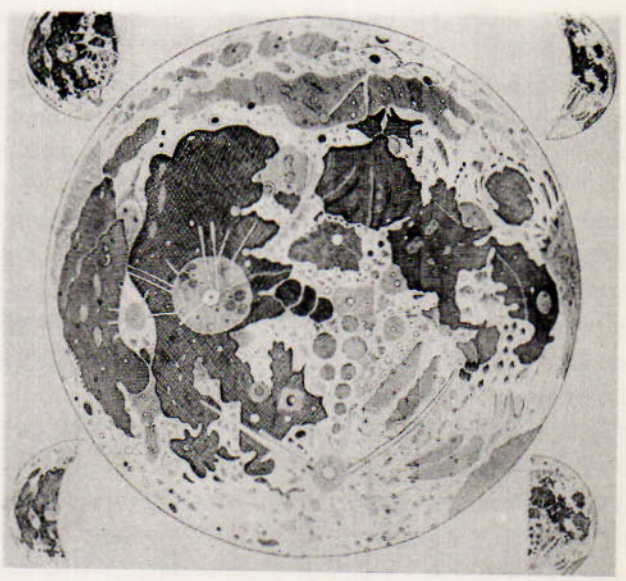


Fig. 242. AN EIGHTEENTH CENTURY CHART OF THE MOON. Engraving from *Introductio ad Veram Astronomiam*, by J. Keill (Oxford, 1718).

Fig. 243 — COMPARATIVE SIZES OF THE MOON AND THE EARTH

according to Kepler's laws, and are therefore not always the same distance apart. The Moon's orbit has a relatively high eccentricity and this causes considerable variation in the apparent size of its disc.

The mean distance of our satellite is 239,000 miles, but owing to the eccentricity of its orbit (0.0549), this may vary from 226,000 miles at perigee to 252,000 miles at apogee. This means a variation in its apparent diameter from 29′ 20″ to 33′ 36″.

The Moon is a very deceptive object. Nearly everyone, especially when seeing a full moon low down in the sky, over-estimates its size. As we said in the first chapter, some people compare it with an orange, a saucer or even a plate. These comparisons are exaggerated and in any case, meaningless if the distance of the object is not stated. Such estimates are, in fact, the result of an illusion, which is well known but not well understood. The best proof of the universality of this illusion is in the way in which a night scene is almost invariably drawn by artists: the Moon is shown much too large in relation to objects in the surrounding landscape.

The only valid way of stating the Moon's size is by giving its angular diameter, as above. For those who are not familiar with celestial measurements, this diameter corresponds to a little disc 1/5 inch in diameter, seen from a distance of about 2 feet. This is easily proved. Cut out such a disc, hold it at arm's length, and it will be found to eclipse the Moon exactly. This is a striking demonstration of the illusions we are subject to when contemplating the 'orb of night'.

Another illusion is the belief that the Moon is larger when on the horizon than when high in the sky. If we think about it we shall see that just the opposite is the case. When the Moon is on the horizon it is further away than when high up, although the difference is infinitesimal, astronomically speaking. It is equal to the Earth's radius, which is 1/60 part of the Moon's distance. Although very small astronomically, this is quite large by ordinary standards. Many people, however, in the course of their lives, do travel much further than the distance from the Earth to the Moon, e. g. sailors and the crews of aircraft. Of course, this is not the same thing as travelling 240,000 miles in a single journey, as shown in Fig. 245. However, the times taken by normal means of travel, supposing they could be applied to the Earth-Moon journey, would not be excessively long.

As usual in astronomy, the distances quoted are between centres and do not allow for the diameters of the bodies in question. The diameters of the Earth and Moon, however, are not negligible compared with their distance apart. Thus from the surface of one to the surface of the other we must deduct 5,043 miles from the separation between their centres. When the Moon is at its nearest, i. e. at perigee, the two bodies are only 220,957 miles apart.

The Moon's motion. — The Moon's mean orbital velocity is 0·63 m.p.s. and it accomplishes a complete revolution in 27d 7h 43m 11.5s; this means that it completes one circuit of the celestial sphere in this time, which is therefore called its sidereal revolution period. It is also the interval between two conjunctions with the same star.

The Moon's speed and distance are such that its movement can easily be observed. Neglecting the diurnal rotation of the sky, which is shared by all the heavenly bodies, it takes only a short time for the observer to detect a change in the Moon's position with respect to some nearby star, as our satellite travels westwards along its path against the background of the constellations. This displacement, to which we shall refer later in our discussion of eclipses and occultations, takes place at a rate of 13° 10′ per day, or approximately 26 times the apparent

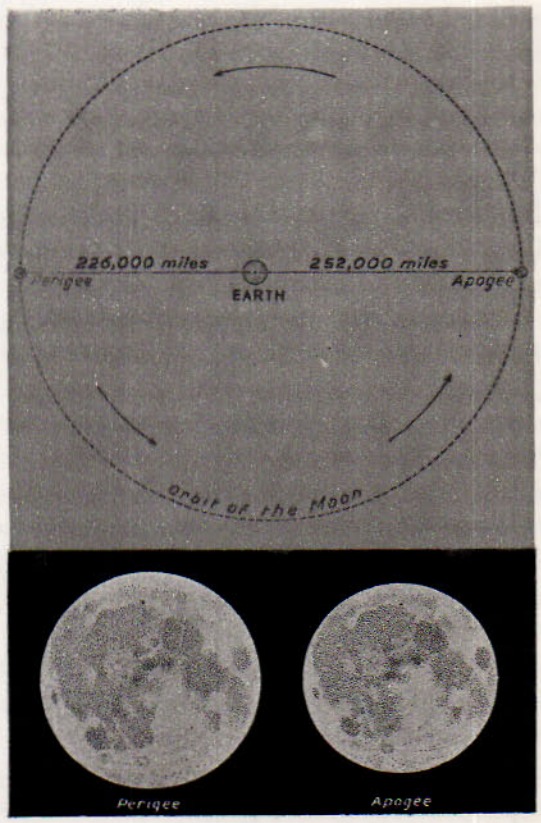


Fig. 244 — ECCENTRICITY OF THE MOON'S ORBIT: apogee and perigee, and the consequent variations in the Moon's apparent diameter.

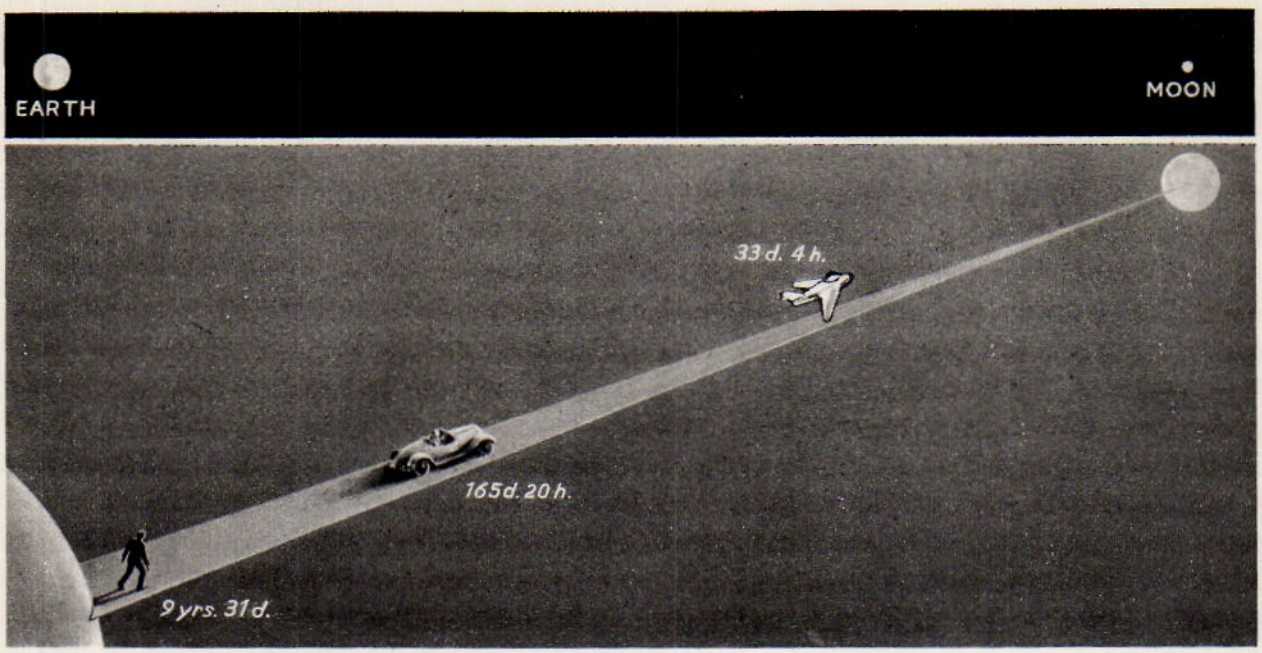


Fig. 245 — THE DISTANCE FROM THE EARTH TO THE MOON. In the upper diagram the distance is represented to scale; in the lower one is shown the time that would be taken to reach the Moon by an aeroplane travelling at 300 m.p.h., a car travelling at 60 m.p.h., and a pedestrian walking at 3 m.p.h.

diameter of the disc. The Moon therefore travels through an angle equal to its own apparent diameter in just under one hour.

The Moon's own motion has the effect of delaying its return to the meridian each day: the mean time between two successive transits is 24h 50m 30s. The mean time is given because there are considerable variations either way; these are due to the eccentricity of its orbit, which means, as we have already shown, that the Moon's velocity is not constant. Further allowance must be made for the inclination of its orbit. This all adds up to a variation in the period between transits of between 38 and 66 minutes.

Moreover, any particular phase of the Moon is repeated only after an interval (called the synodic period) two days longer than the sidereal revolution period mentioned above. The causes of this difference are easily understood. The reason is, in theory, identical with the one for the difference between the sidereal and the solar day. Let us therefore recapitulate this explanation and apply it to the case of our satellite, as shown in Fig. 247. Starting at conjunction (A) with the Sun (the phase known as new moon, which is used to measure the start of the synodic period) we see that, as the Moon goes round the Sun in company with the Earth, our satellite makes one complete sidereal revolution to A′ and must then travel a little extra distance (to B) in order to be once more exactly between the Sun and Earth.

The mean value — since irregularities in its velocity also make themselves felt — is 29d 12h 44m 3s; this is called a lunation or lunar month. The diagrams accompanying this text should only be taken as intending to

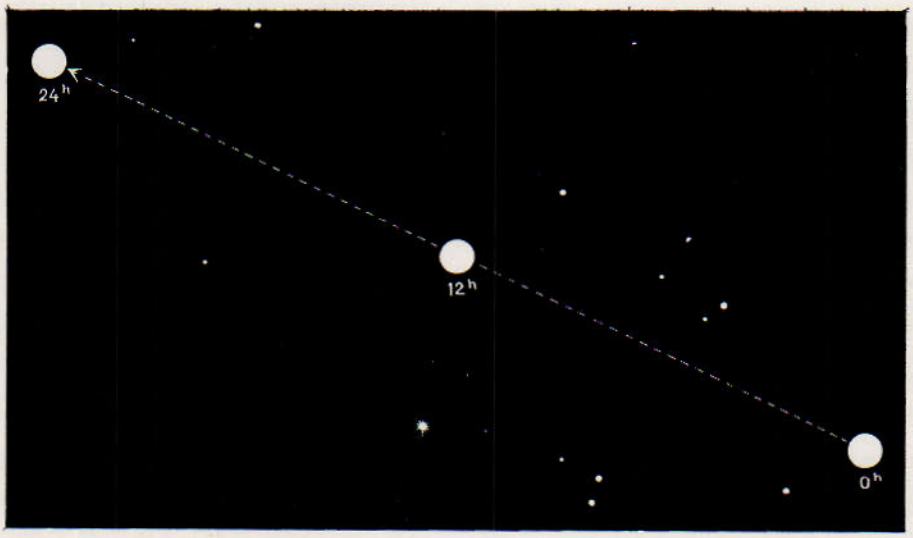


Fig. 246 — APPARENT MOTION OF THE MOON AGAINST THE STARS IN A PERIOD OF 24 HOURS

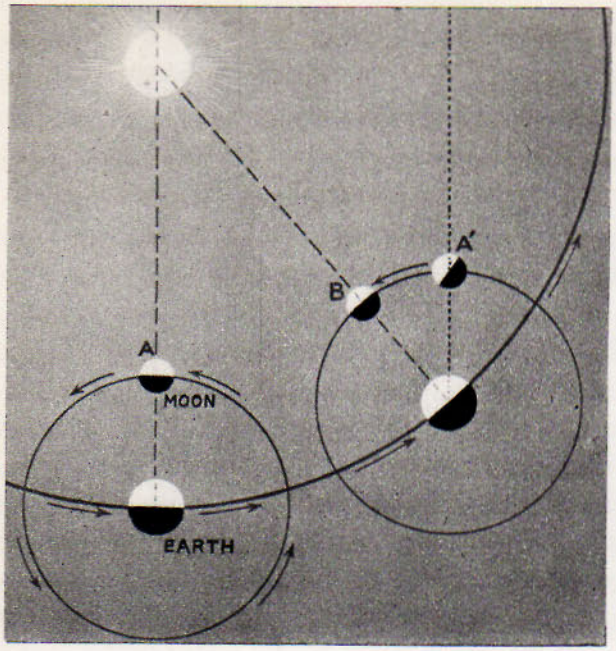


Fig. 247 — SIDEREAL AND SYNODIC REVOLUTION

illustrate this principle as simply as possible, and they do not give a correct impression of how the Moon actually moves through space. This is because the Moon's path is a very sinuous curve, a combination of its orbit round the Earth and the Earth's orbit round the Sun. The picture is further complicated by the inclination of the Moon's orbit and the many smaller irregularities discussed in an earlier section.

The Moon's phases and the earthlight. — Like all planetary bodies, the Moon has no light of its own: it shines simply because it is illuminated by the Sun. We may recall what has already been said about the illumination of a spherical globe in space: light can fall on only one half, while the other remains dark.

With this in mind, the principle governing the Moon's phases may be grasped without difficulty. The time in which they occur is governed by the synodic period; any particular phase depends on the Moon's position relative to the Sun which illumines it and the Earth from which it is seen.

In Fig. 250, which has been simplified, the Earth is supposed to be at rest with respect to the Sun, whilst the Moon travels in a circular orbit in the plane of the ecliptic. It will be noticed at first glance that the aspect under which we on the Earth see the illuminated portion of the Moon's disc (which is always facing the Sun) varies at different points around the Moon's orbit.

At new moon, the whole of the dark half is pointing in our direction, and the Moon is therefore invisible. Then, as the Moon begins to move round its orbit, making a greater and greater angle with the line joining the Sun to the Earth, more and more of the illuminated portion becomes visible from the Earth, and through effects of perspective on the Moon's globe, we see it in the form of a crescent. This increases in width as the angle widens. When the Moon is at right angles to the Sun, exactly half of the daylight hemisphere can be seen: this phase is called the first quarter. From now on, the illuminated half is turned more and more directly towards us, the darkened part disappearing at the same rate as a rapidly narrowing crescent of darkness (which, of course, we cannot see but only infer). When the Moon is exactly opposite the Sun, the illuminated hemisphere is pointing exactly towards us and we see a fully illuminated disc — hence the name full moon.

The phases which follow the full moon are produced in exactly the same way but in the reverse order, finally returning to new moon, when the cycle begins over again.

The relative positions of the Earth and Moon explain not only the phases but also certain effects which are sometimes no less remarkable. There is, in particular, the Earthlight, or as it is more commonly known, the Old Moon in the Young Moon's arms.

When the Moon is only a thin crescent just before or after new moon, we can often make out the rest of the disc faintly illuminated, completing the circle of the crescent. This ghostly disc is in fact the dark side of the Moon, and should, if what we said earlier is strictly true, be invisible at this time. But, as will be seen from Fig.250, the Earth is illuminated by the Sun in exactly the same way as the Moon; and, as seen from the Moon, it would therefore go through a corresponding set of phases. These phases are exactly the opposite of the Moon's phases as seen from the Earth: when the Moon is at first quarter for us, we are at last quarter for the Moon, and when we are at new moon, the Moon is at 'full Earth'.

In this respect, the Earth may be thought of as the Moon's satellite, and the faint illumination is nothing more than 'Earthlight' shining on the night side of the Moon which is turned towards us. Earthlight is much stronger than moonlight for the simple reason that the Earth is 13 times larger when seen from the Moon than the Moon as seen from Earth, and is also a much better reflector of light than the Moon.

The Earthlight is most in evidence a few days either side of new moon, but fades considerably as the Moon's phase increases, since the Earth's phase is decreasing at the same rate. Earthlight is also less easily seen when the lunar phase is large, since the eye is dazzled by the

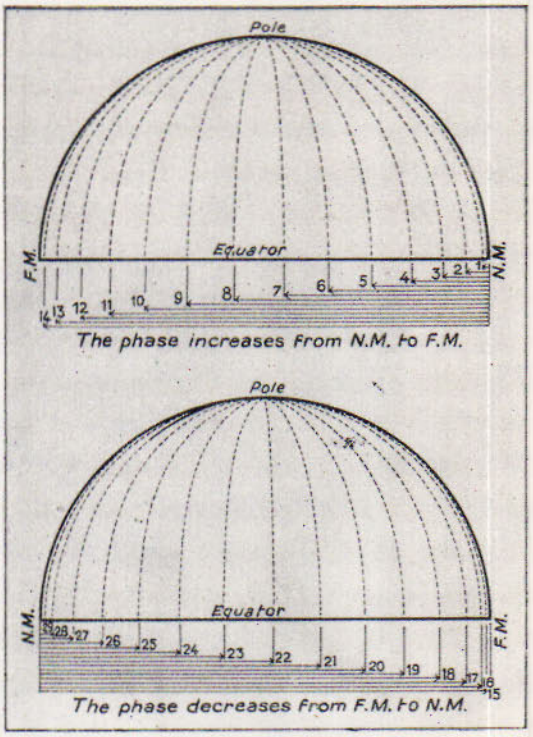


Fig. 248 — PROGRESSION OF THE TERMINATOR ACROSS THE MOON'S DISC DURING THE COURSE OF ONE LUNATION. The figures indicate the days of the lunation.

Fig. 249 — THE PROGRESSIVE ILLUMINATION OF THE LUNAR HEMISPHERE TURNED TOWARDS THE EARTH

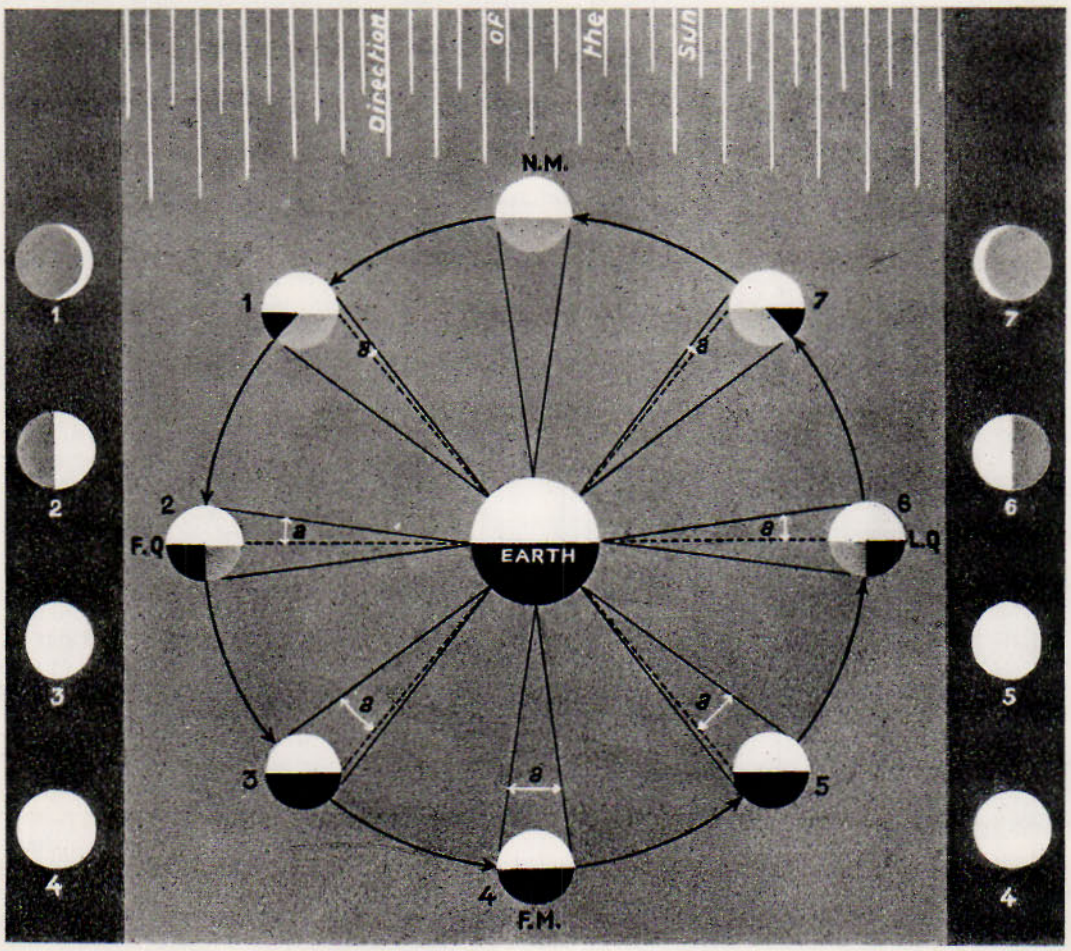


Fig. 250 — THE PHASES OF THE MOON AND THE REASON FOR THE EARTHLIGHT. As the Moon moves round the Earth (1, 2, 3, 4), the illuminated hemisphere turns more and more towards the Earth (angle *a*); afterwards (5, 6, 7), the process is reversed. Seen from the Moon, the Earth exhibits exactly the same phases, but the other way round; at new moon, the Earth is full and casts a considerable amount of light back on to the dark side of the Moon, which is then illuminated by the Earthlight.

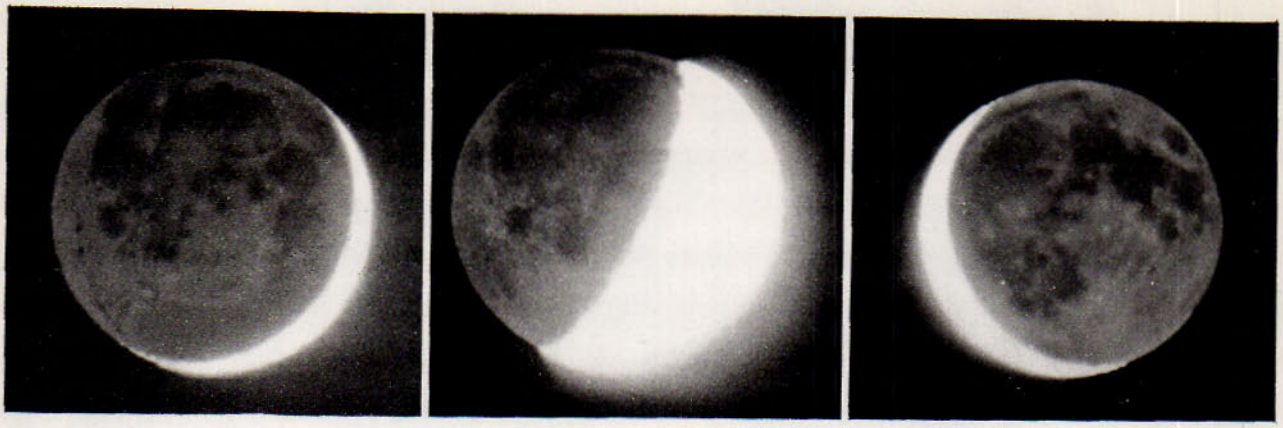


Fig. 251 — THE EARTHLIGHT PHOTOGRAPHED AT DIFFERENT PHASES OF THE MOON. The illuminated portion of the Moon is highly over-exposed.

brightness of the illuminated portion. With the naked eye, therefore, it is clearly seen only when the Moon is a thin crescent, but in the telescope the Earthlight remains visible right up to just before full moon, when it is very faint indeed.

The Moon's rotation. — Like all astronomical bodies, the Moon rotates on its axis. In this case, the rotation period is equal to the Moon's revolution period round the Earth, which means that our satellite always turns the same face towards us. This is a very widely known fact, since the same pattern of markings — the 'face' of the so-called Man in the Moon — is always visible. This frequently gives people the erroneous idea that the Moon 'does not turn'.

Why this should be so is easy to understand. In Fig. 254, the Moon is represented as travelling round a circle with the Earth at the centre (E). Starting from the new moon, take the point A on the opposite side of the Moon to the Sun, and therefore facing the Earth. At first quarter, the Moon has completed a quarter of its journey. If were not rotating (that is to say, if it maintained an absolutely fixed orientation in space), A would now be at A_1; but since the rotation period equals the revolution period, the Moon has in fact made one quarter-turn, in the direction of the arrow, from A_1, which shows the initial direction: thus the point in question is still facing the Earth. The same situation is repeated at full moon, and so on.

Fig. 252 — THE START OF A LUNATION: the slender crescent of the Moon together with the Earthlight photographed 34 hours after new moon.

A simple experiment will vividly demonstrate what happens. Walk round a circular table, always facing the centre of the table. When you return to your starting point, you will see that you have been facing the table all the time so that from a point in the centre only your face was visible, not the back of your head. Nevertheless, you yourself will have made one complete turn 'on your axis', just as if you had stayed where you were and pivoted round. Returning to the Moon, the diagram also shows that every part of one lunar hemisphere will be seen from Earth under every possible condition of illumination.

Let us continue our explanation with reference to point A, Fig. 254. From new moon onwards, this portion of the Moon will only be visible by the reflected light of the Earth, and will not emerge into daylight until first quarter. At this time the Sun is rising at this particular spot; the rays of light just catch the surface horizontally and every tiny irregularity casts a long black shadow. This is best shown on a model. As the terminator moves on from there, the direction of light becomes more nearly vertical. The Sun is directly overhead at full moon, when all shadow effects disappear and the landscape takes on a flat, monotonous appearance. This sequence of events is repeated in the reverse order during the second half of the lunation.

Thus any part of the Moon's surface, although continually pointing in our direction, will be seen under all conditions of illumination during the course of a month.

The above explanation is designed merely to give a broad outline of the mechanism causing the Moon to keep the same face turned in our direction. As usual, what actually happens is far more complex. The two movements — rotation on its axis and revolution round the Earth — do not coincide so exactly as our description might seem to suggest. The reason is that although the two periods are equal, the rotation takes place uniformly, whilst the Moon does not revolve round the Earth at uniform velocity owing to the fact that its orbit is an ellipse. If we employ an imaginary mean period as in our description, slight variations will occur in either direction. Before going into this matter further, we must note a few other features of the Moon's behaviour with respect to the Earth.

The fact that the same side of the Moon always points in our direction appears to be the result of tidal effects

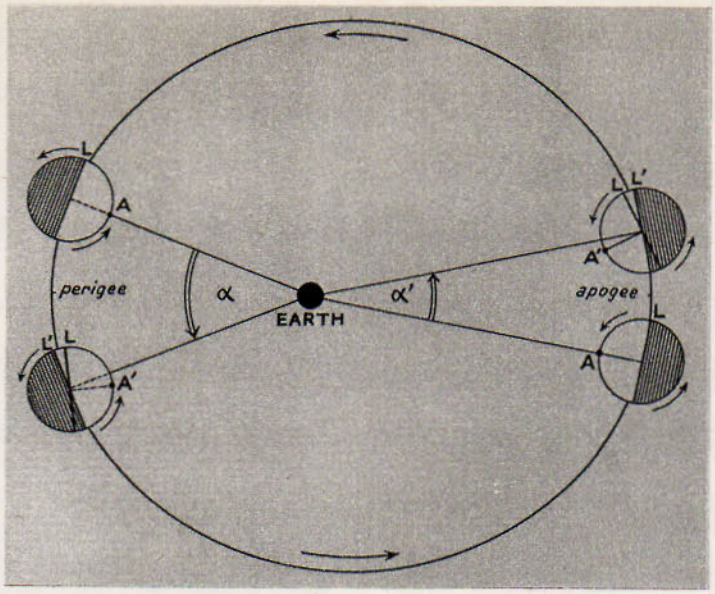


Fig. 253 — THE PRINCIPLE OF LIBRATION IN LONGITUDE

between the two bodies at an early period in their history, when our satellite was still in plastic condition. These effects consisted of a deformation of the Moon's globe, and will be better understood when we come to study the tides. In short, though, they led to a sort of braking through firction which tended to slow down the rotation, as we have seen occurring in the case of the Earth itself.

In the Moon's case, the result would be to pull it into the shape of an egg, with the large end pointing towards the Earth.

This is a purely theoretical matter since the amount of deformation is not much more a few hundred feet at the most, which is very small indeed compared with the Moon's diameter of 2,160 miles. In fact it is well below the height of many surface features such as mountains, crater rims, etc.

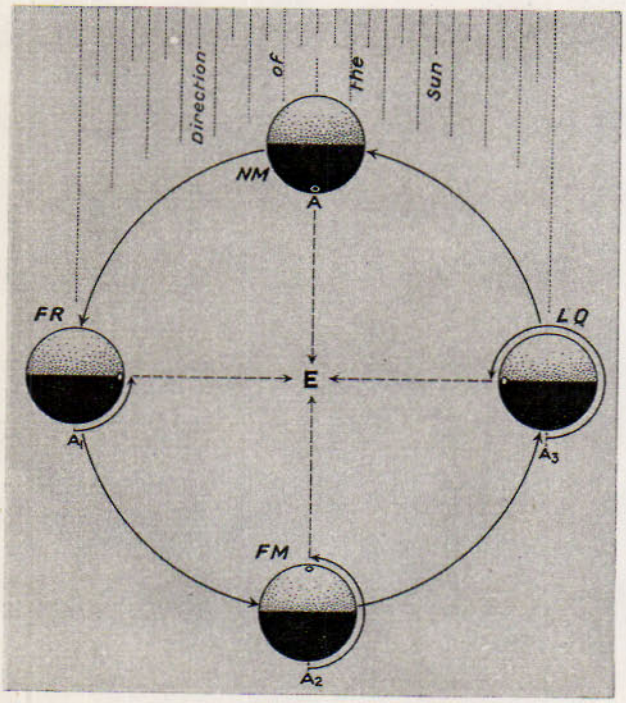


Fig. 254 — THE COMBINATION OF REVOLUTION AND ROTATION WHICH KEEPS ONE FACE OF THE MOON TURNED PERMANENTLY EARTHWARDS

Returning to the question of the Moon's motion, the elliptical nature of its orbit means that its velocity will be greatest at perigee and least at apogee. Taking these two extremes for simplicity, we see that (Fig. 253) the Moon travels through an angular distance α and α' in equal times, and that α is greater than α'. Therefore the revolution will not match the rotation, which is uniform. This means that a point A, exactly on the line joining the centres of the Earth and Moon, will 'gain' or 'lose' with respect to that line according to whether the Moon's angular displacement is smaller or greater than the angle through which the lunar globe has turned in the same time. This lag will, of course, affect the whole surface equally so that features on the edge of the disc, for example, will not always look the same: sometimes some will disappear at one edge whilst others become visible at the opposite side.

And so to us the Moon will seem to move with a sort of rocking movement either side of a mean position. This is called libration, the period of which is half-lunation or about 14 days.

As this rocking motion is in the plane of the orbit, it is called libration in longitude, but there are other librations which further complicate matters. The Moon's axis is slightly inclined, with respect to the plane of the orbit, by an angle varying between 83° 11′ and 83° 29′; consequently each of the lunar poles tilts in our direction in turn. This also occurs in a period of 14 days and is called libration in latitude. It has the effect of bringing into view parts of the polar regions which are not normally visible; but at the opposite pole regions normally exposed then move over the edge of the disc.

There is yet another effect to consider. The Moon is near enough and the Earth is large enough for the terrestrial observer's position on the globe to make some difference to how he sees the Moon. In Fig. 255, if the observer moves from A to B because of the Earth's rotation, the part of the Moon visible to him will change from $a' - a'$ to $b' - b'$. This is called diurnal parallax.

On account of these different factors, therefore, the surface features are continually changing position as

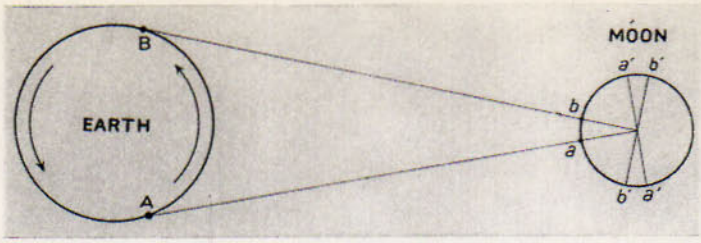


Fig. 255 — THE PRINCIPLE OF DIURNAL PARALLAX

a whole, and those at the edge or limb of the Moon alternately vanish and reappear. Libration does not correspond to the phases, which means that an object just on the terminator at first quarter, for example, will not be found there at exactly the same phase of the next lunation. Libration means in fact that half of the Moon is always visible, and that a certain belt at the edge of this hemisphere is sometimes invisible and sometimes visible whilst the rest of the Moon is permanently hidden. The width of the intermediate zone is about 8°.

Owing to the libration in latitude, there are two zones round the poles, each 6° 50′ in width, which alternately appear and disappear, whilst the width of the diurnal libration zone is about 1° 2′. The final effect of all these factors (see Fig. 258) is that instead of half the Moon we can observe about 59% of the total surface area.

Inclination and displacement of the Moon's Orbit. — Up to the present we have considered the Moon's orbit separately, but now we must examine it together with the Earth and see how its orientation in space determines the way in which the Moon moves in relation to the Earth.

The plane of the Moon's orbit is inclined at an angle of 5° 8′ to the plane of the ecliptic (which is the Earth's orbit). This means that, as seen from the Earth, the Moon travels round the Earth along a circle which makes an angle on the celestial sphere of 5° 8′ to the ecliptic, the circle along which the Sun seems to move in its yearly

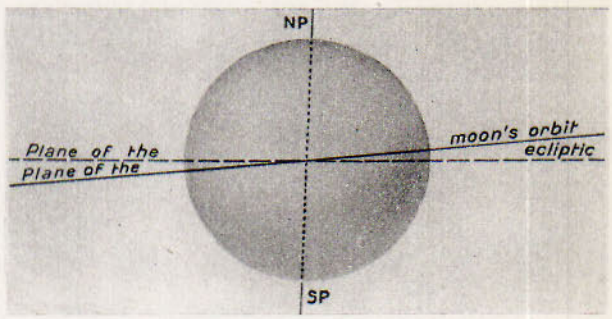


Fig. 256 — INCLINATION OF THE MOON'S AXIS

journey. These two circles cut at two points NN′ (Fig. 259), the line of intersection of the two planes, called nodes. Thus the Moon's monthly circuit of the sky is half above and half below the ecliptic, which it crosses alternately at the two points. This path is important in connection with the phases.

For simplicity in earlier explanations, we imagined the orbits of the Earth and Moon as lying in the same plane, in which the Sun also would be situated; but this was merely a device and as the reader will no doubt already have realised, the inclination of the moon's axis means that the actual relative positions of the three bodies at any particular phase will be somewhat different from that suggested by these simplified explanations.

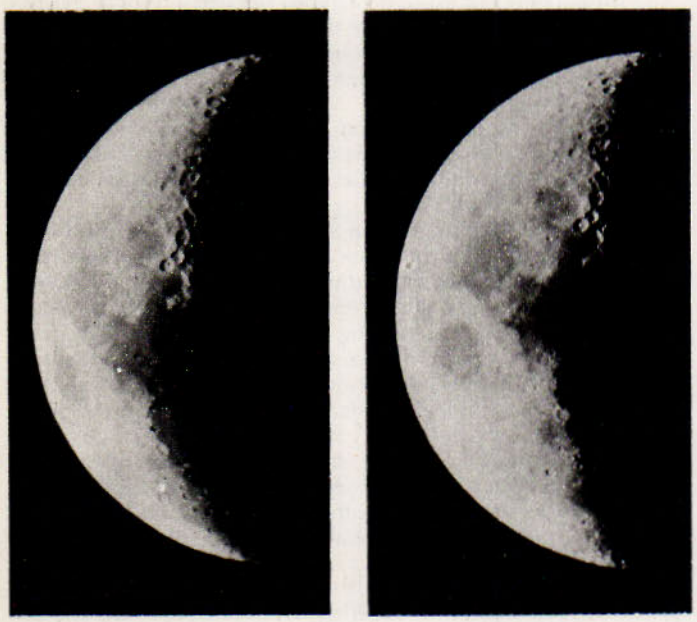


Fig. 257 — EFFECTS OF LIBRATION. Objects near the limb may become very foreshortened, while the same objects will not always appear on the terminator, although the phase may be the same, if the libration is different.

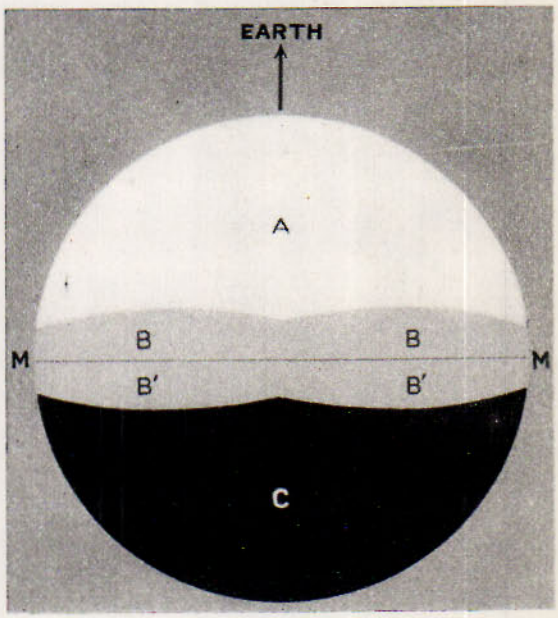


Fig. 258. THE EFFECT OF LIBRATION ON THE VISIBILITY OF THE MOON'S GLOBE. A is always visible; B and B′ are alternately visible and invisible either side of M; C is always invisible.

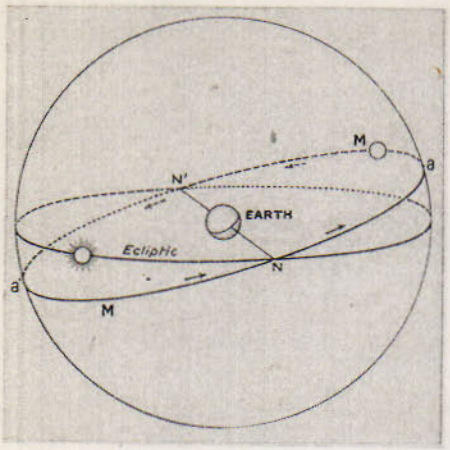


Fig. 259 — APPARENT MOTION OF THE MOON WITH RESPECT TO THE ECLIPTIC. N, ascending node; N′, descending node.

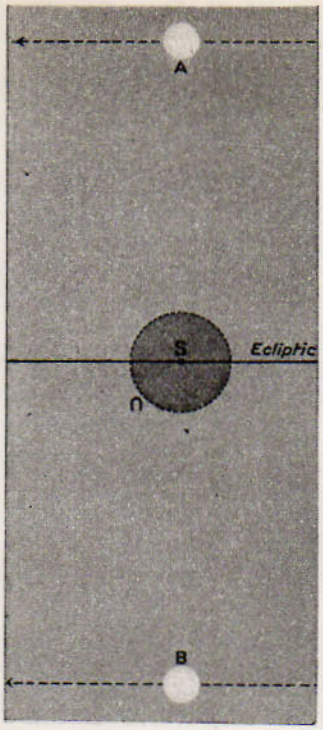


Fig. 260 — MAXIMUM DISTANCE BETWEEN THE FULL MOON AND THE EARTH'S CONE OF SHADOW (U), THE CENTRE OF WHICH (S) IS EXACTLY OPPOSITE THE SUN

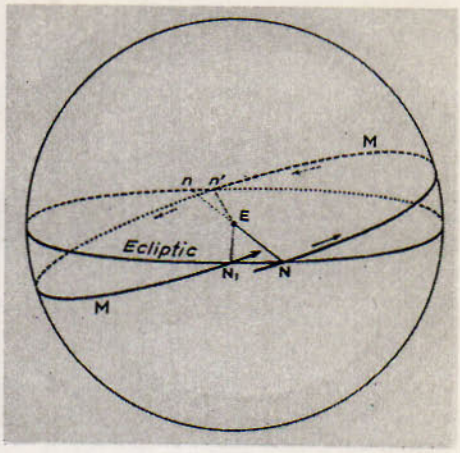


Fig. 261 — RETROGRADE SHIFT OF THE NODES OF THE LUNAR ORBIT. N, N′, ascending nodes; *n*, *n′*, descending nodes.

According to these, at new moon our satellite should be exactly on the line joining the centre of the Earth to the centre of the Sun, and between these two; at full moon it should be on the same line, but on the opposite side of the Earth from the Sun. If this were the case, at every new moon our satellite would pass between the Sun and Earth, causing an eclipse of the Sun, whilst at full moon it would itself be eclipsed as it passed into the Earth's shadow. This is the basic mechanism of eclipses, as will be shown later.

But we all know that eclipses occur only at fairly long intervals. There are two main reasons for this. First, there is the inclination of the Moon's orbit; second, the displacement of the plane of the lunar orbit in space. In order to see what the latter type of movement means, let us recall the phenomenon of precession, the top-like wobble of the Earth. The Moon's orbit behaves in a rather similar way, so that the line of the nodes actually revolves in space, making one complete turn (in the retrograde direction) in 18 years.

Let us now look at Fig. 259. The Moon's path (M) coincides with the ecliptic only at the two points where they intersect: the ascending node N and the descending node N′. From there on they become further and further apart up to the points *aa′*, half way between the nodes. This distance is exactly equal to the inclination of the lunar orbit, i. e. 5° 8′. Thus different phases may occur at varying angles from the ecliptic, to which the Sun always keeps: at certain times the Moon will be just in this plane, whilst at others it may be above or below it. For an observer on the Earth, this is seen as a slight change in the visible part of the Moon's disc.

The conditions undergo a regular change owing to the retrograde movement of the nodes as shown in Fig. 261, this diagram being complementary to Fig. 259. At each circuit, therefore, the Moon crosses the ecliptic a little further on than where it would if its orbit were stationary. Let us follow its progress from the ascending node N. During one half-revolution, the slow movement of the orbit itself causes the descending node to drift from *n* (exactly opposite to N) to *n′*. Then as the Moon comes round the second half of its orbit, it re-cuts the ecliptic at N_1, the new position that the node has reached. This process continues, the nodes moving further and further round the orbit until they return to the starting point. After 18 years, therefore, the Moon's orbit will have passed through every possible orientation in space with respect to the Earth.

This is one of the reasons for the infrequence of lunar and solar eclipses, which can occur only when the Moon is passing through one of the nodes, the latter having to be directly in line with the Sun and Earth. The frequency of eclipses is therefore determined by the interplay between the revolution of the Earth and the retrogradation of the nodes of the Moon's orbit. Except at certain times, which will be seen from the diagrams, new moon occurs outside the plane of the ecliptic, which means that the Moon will pass either above or below the Sun as seen in the sky, their maximum separation being approximately 5°. In the same way, we can still see the Moon at full because it passes above or below the shadow cast by the Earth. Strictly speaking it is not correct to call this 'full moon'; as our satellite may be as much as 5 degrees away from the true position, the observer's line of sight makes this same angle with the direction from which the the illumination is coming, and the disc appears to be slightly tilted form north to south. This produces a slight phase effect round one of the poles, making the relief stand out sharp and clear.

Naturally, between these two extremes there is every

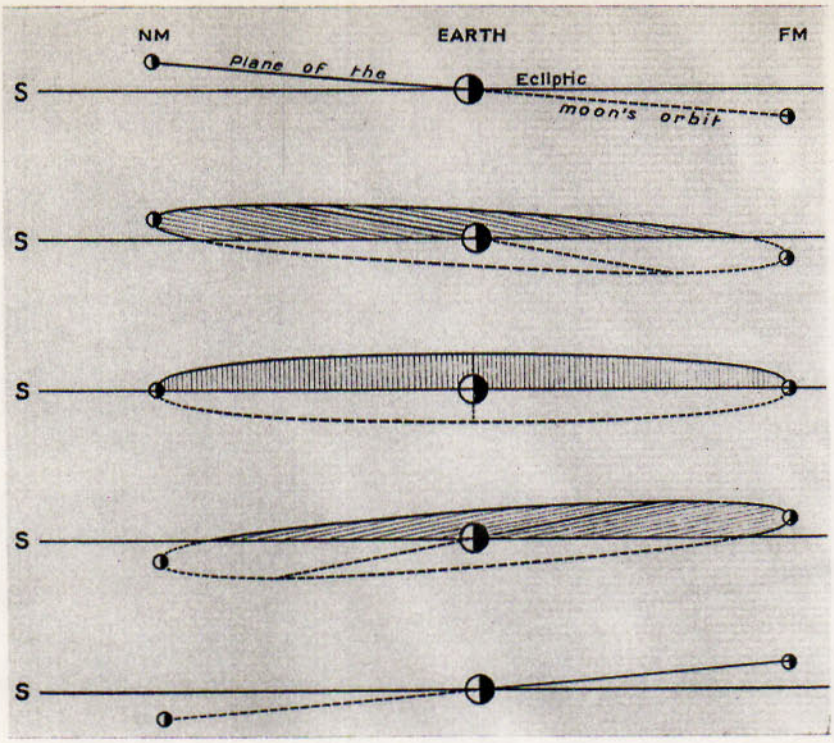


Fig. 262 — THE INCLINATION AND VARIABLE ORIENTATION OF THE PLANE OF THE MOON'S ORBIT ON THE PLANE OF THE ECLIPTIC. The Moon's orbit is shown in perspective, the shading denoting the half above the ecliptic.

intermediate stage. This does not apply only to the new and full moon: every other phase may occur with a similar gap between the Moon and the plane of the ecliptic.

In addition to this, we must consider the effect of the eccentricity of the lunar orbit, which implies different distances from the Earth and different velocities while it is passing from one position to the other.

Furthermore, the orbit turns in its own plane with what may be considered uniform motion in the direct sense. The major axis completes one turn in 9 years, and this is called the progression of the line of the apsides.

We shall not attempt to give a complete account of the Moon's motion in these pages; to deal with the matter thoroughly would be beyond the scope of this work, and in any case, it may be said in passing, to do the job properly would take the whole lifetime of a mathematician . . . We may, however, note that there are two further important irregularities to be considered. They are:

Evection, with a period of 31½ days, which may displace the Moon 1° 16′ from its true position;

The Annual Equation, with a period of one year and an amplitude of 11 minutes of arc.

Thus our satellite steers a very unsteady course through space. The consequences of this are important for lunar observers, for it means that the same conditions are never repeated exactly from one lunation to the next. It is only after a period of 18 years that the cycle begins anew, this being, broadly speaking, the time taken by the nodes to make a complete circle of the ecliptic.

All this, which we shall discuss further with respect to eclipses, is of importance in the study of the Moon's surface, since it determines the exact conditions of illumination under which it is seen, and these variations in illumination may well be mistaken for real physical changes by an inexperienced observer.

The visibility of the Moon. — In this section we shall discuss the apparent motion of the Moon, including questions such as why its path across the sky, as seen from any particular spot, is sometimes low down near the horizon or high up, perhaps almost overhead, during the course of one lunation, and why this situation changes from season to season.

These variations are due to much the same causes as the Sun's variation in altitude, as explained in an earlier section. It may be applied, with certain modifications, to the Moon's motion, which takes place very nearly in the plane of the ecliptic, although the fact that the two do not exactly coincide means that the variations are exaggerated. Indeed at any particular place the Moon may be 5° above or below the Sun's position at a corresponding time; at the solstices, it may be remembered, the Sun's celestial latitude is 23° 27′ N. or S.

When the ascending node of the lunar orbit coincides with the vernal equinox (the point where the Earth's equator intersects the ecliptic) the angle between the plane of the lunar orbit and the equator is at its maximum, and the Moon reaches a celestial latitude of 23° 27′ + 5° 8′ which equals 28° 35′ N. or S. When the descending node coincides with the vernal equinox, this angle is at its minimum and the Moon only reaches 23° 27′. − 5° 8′, which is 18° 19′. Thus its height in the sky varies between these two extremes, which is to say an over angular distance of 57° 10′ in the first case and 36° 28′ in the second.

If, therefore, we know how to relate these data to the horizon of the place, we can then find out at what height in the sky the Moon crosses the meridian. The calculation is elementary and follows from the explanations given earlier about the appearance of the celestial sphere at different times, as a function of the observer's position on the surface of the Earth.

We shall consider what are the conditions at one

Fig. 263 — THE FULL MOON, SHOWING A SLIGHT PHASE EFFECT ROUND THE NORTH POLE. In this photograph, as in the usual astronomical practice, north is at the bottom and south at the top, to conform with the way objects are always seen through a telescope.

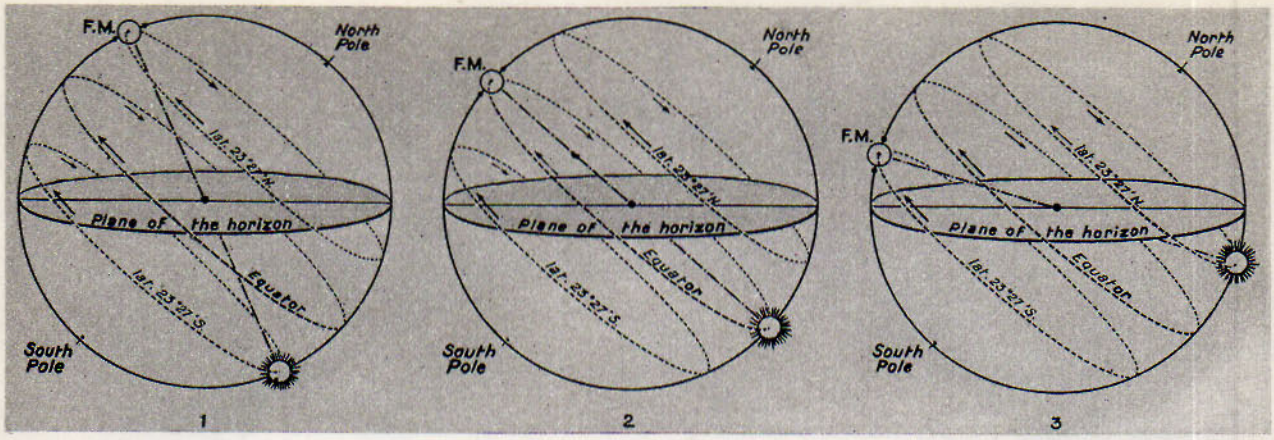


Fig. 264 — THE MOON'S APPARENT MOVEMENT IN THE SKY FOR A TEMPERATE LATITUDE IN THE NORTHERN HEMISPHERE. The height of the Sun and Moon at mid-day varies considerably at different times of the year: 1, the winter solstice; 2, the equinoxes; 3, the summer solstice

particular place — the latitude of England for example — paying special attention to the Moon's relation to the ecliptic. Even without going into figures, it is possible to realise the affinity between the behaviour of the Moon and the Sun; the latter's seasonal variations are the same at those of the Moon, but occur in the reverse order, the full moon, for example, being opposite the Sun. If we in our northern temperate latitudes find that the Sun remains low on the horizon all day in winter because of its strong southern declination, the Moon is in exactly the opposite position, as high in the sky as the Sun appears to be in summer. On the other hand, in summer the Moon occupies the Sun's winter position and is always affected in some degree by the mist and dust which obscure the sky in the vicinity of the horizon. Similar differences in the phases occur at other times of the year, the Moon always occupying the position which the Sun will occupy six months later. Thus in spring (in northern latitudes) the first quarter is high in the sky, and the phenomenon of the Earthlight is observed to best advantage. In autumn, on the other hand, conditions are unfavourable, the Moon's path lying too near the horizon.

Just as for the Sun, this sequence reverses as we pass from the northern hemisphere to the southern. In the polar regions, moreover, the Moon remains above the horizon for a period which may be as long as 14 days, or half of one revolution, thus restoring some light from the hidden Sun at times when the latter is invisible for weeks or months on end. Indeed, we have only to read the experiences of polar explorers to realise how welcome this light may be. Captain O. Sverdrup, who spent four consecutive winters exploring the then unknown North American archipelago, writes: 'This absolutely limpid moonlight is quite perfect. For us poor hermits of the ice, it is our only day. Bathed in its beautiful rays, the landscape shimmers with an almost supernatural radiance, full of the most astonishingly sharp contrasts: the ice and snow are of an extraordinary whiteness, the shadows are of such a deep black that words cannot do them justice...'

From this eulogy, however, it should not be thought that in these regions moonlight is more intense than elsewhere: its brightness is simply enhanced by the nature of the landscape, which matches the quality of moonlight so well. Moreover, the human eye is specially glad of any light relieving a long period of darkness, and tends to endow it with more than its true intensity. In view of this, therefore, let us pause for a few moments to discuss the whole question of nocturnal illumination of the Earth by the Moon.

The intensity of moonlight. — It is not enough to know simply that the Moon reflects the light of the Sun back on to the Earth's surface; we must know what proportion of the Sun's radiation comes to us in this way.

The Moon is not quite comparable with a mirror. Its surface is very varied and as a whole is a poor reflectore of the light which falls on it.

As everyone has probably noticed, the intensity of moonlight is very variable. It increases from the start of the lunation onwards, reaches its maximum value at full moon and then decreases again. Thus the brightness of the Moon is determined principally by the phase, that is to say the amount of surface area available to reflect the light.

There are, however, several more factors which affect the final intensity of the reflected light; in other words it is not entirely determined by the available surface area. For instance, the part of the Moon we see at first quarter consists of half of the disc visible at full moon, but it must not be concluded on this account that the illumination at first quarter is half that of the full moon, as would result from covering and then uncovering half a flat, illuminated disc of cardboard. Measurement shows that the intensity of the illumination given by the first quarter is only one-twelfth that of full. The reason for this is as follows:

On a solid sphere, the illumination of equal surface areas decreases as the obliqueness of the rays of light increases. For this reason, the lunar globe exhibits a progressive fall-off in brightness towards the boundary between night and day, or the 'terminator', as it is called. To this must be added a second factor, which comes into play at the quarters, especially because of the way in which this part of the Moon is then seen. Where the Sun's rays just graze the surface at a tangent, the broken nature of the lunar relief becomes important, since any piece of high ground — a mountain, a crater-rim, etc. — casts a long shadow. The total area of all the shadows must be deducted from the total visible area of the Moon available for reflection, with a consequent decrease in the general intensity of the moonlight. When the Moon is in the crescent phase, the parts then visible happen to be those which, in proportion, reflect light least well, thus causing a further decrease in the general brightness.

Moreover, not all parts of the surface are equally good reflectors of light, the poorest being those large grey areas of *mare* which cover a large proportion of the visible hemisphere. The amount of grey-toned ground visible depends on the phase; as may be seen from

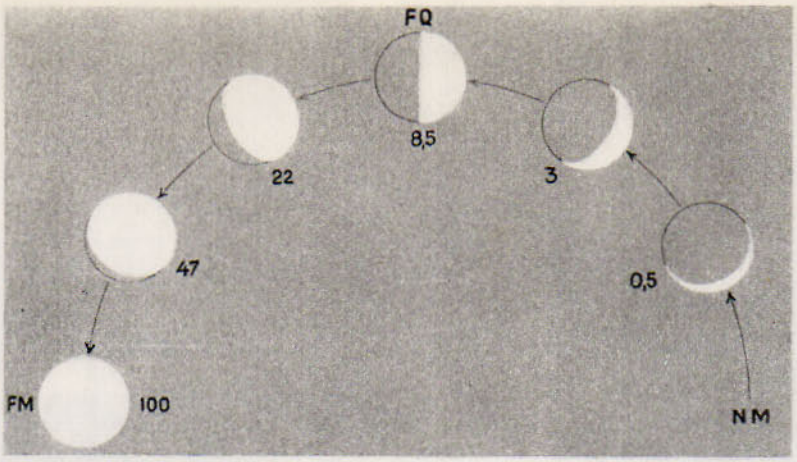


Fig. 265 — INTENSITY OF THE LIGHT GIVEN BY THE MOON AS A FUNCTION OF ITS PHASE

Fig. 268, at or near the last quarter, the dark areas preponderate, rendering this the least bright of the phases.

These differences may, of course, be somewhat affected by the variations in the distance separating the Moon from the Earth, but even more important is the Moon's variable height above the horizon, which introduces atmospheric absorption in proportion to the transparency of the air. Useful comparisons can therefore be made only under conditions which are maintained as uniform as possible.

We may compare the illumination given by the various phases, taking full moon as 100. This is shown in Fig.265, where the phases selected follow at 30° intervals from the Sun. For the sake of simplicity, the diagram deals only with the first half of the lunation; it will show the second if reversed, remembering that the phases after full moon are slightly less brilliant than those before. The most striking feature is the rapidity with which the illumination falls off from one phase to the next.

In view of this, it may be asked what is the maximum effective illumination received when the full moon is seen high in a dark sky on a clear night. This has been measured exactly, giving the following result: the illumination received on a plate at right angles to the direction from which the light is coming, is 0.25 *lux*, i. e one-quarter of that from an international standard candle at a distance of one metre. Or in more familiar terms, moonlight is about as strong as the light given by a 40 watt bulb at a distance of 15 yards or so.

This shows that the Moon is very parsimonious in the amount of light it reflects back from the Sun. Exactly how mean our Queen of Night really is may be gathered from the figures: even at its best, the Moon gives only 1/450,000 of the Sun's light. Strange to say, however, this is not at all the impression we have from contemplating that sort of moonlit landscape which is the joy of poets; this fact gives elequent testimony to the accommodation power of the human retina, which is so sensitive and adaptable that we can see perfectly well by moonlight, after a few moments to get used to the dark.

This, however, does not prevent our gaining a false impression as to the colour and quality of the Moon's light. The surface of our satellite is not a very good mirror — which is demonstrated by the above figures. Indeed, it absorbs about 93% of the incident light. Moreover, the small percentage which is reflected eartwards undergoes a certain change; it loses some of its short wavelengths: that is to say it becomes redder in colour than sunlight.

This fact would appear to be at variance with what we actually see: artists, for example, use bluish tints when painting a moonlit scene, and these colours are often employed in the theatre to give the same effect. This raises the interesting question of why we endow moonlight with a quality which it does not possess, at least in such a degree.

This fact — that the Sun is bluer in colour than the Moon but gives us the opposite impression — is an illusion known as Purkinje's phenomenon. The mechanism is as follows: the eye becomes progressively more sensitive to radiation of short wavelength — blue, violet — as the source of light decreases in intensity. Thus, because of its faintness, and despite its comparative richness in reds and yellows, the Moon's light will give us the impression of containing more blue wavelengths than it actually does, whilst the extreme brilliance of the Sun will emphasise the reds and yellow end of the spectrum to the detriment of the blue.

The above explanation applies only to visual observation. In photography the contrast is even greater, since moonlight is 650,000 times weaker than sunlight.

However, the faintness of moonlight is no obstacle to photographers who wish to try their skill, providing they do not ask of it more than it can do. Since it is always necessary to give a long exposure to photographs taken by moonlight, there are in practice greater limits to photography by moonlight than to photography by sunlight. By this we mean that even with present-day emulsions we cannot take 'snapshots' by moonlight, although it is nevertheless possible to obtain good pictures with fairly short exposures, thus obtaining some very useful documentary records.

It goes without saying that the widest aperture and the fastest plates should be used. Panchromatic plates are best adapted to the nature of moonlight. We must not forget to allow for the phase of the Moon, its height above the horizon and the transparency of the atmosphere. Since so many factors must be taken into account and since it is impossible to give any useful information about many of them, the best road to success in this field is the one of trial and error. Remember that full moon is the best time in theory, although the Moon's low altitude in the summer may well mean that better results will be obtained when the phase is less but the altitude is greater. One very quickly begins to learn how to assess all these factors.

Generally speaking, landscapes may be successfully photographed with perfectly ordinary cameras, providing

Fig. 266 — MONT ST MICHEL (FRANCE) PHOTOGRAPHED WITH THE MOON RISING TOWARDS THE END OF THE TWILIGHT. Exposure 1 minute with yellow filter.

sufficient exposure is given. This will usually be several minutes, at least. Although stationary objects such as trees, buildings, hills and the like may be photographed in this way, it is much more difficult to catch moving or fleeting objects such as clouds and other things in the sky. This is unfortunate, since it is very interesting and valuable to continue recording such phenomena as clouds during the hours of night, as an extension of our daytime observations.

Sky and clouds are always more 'photogenic' than details of the landscape, white objects excepted. For this reason, as most amateur photographers will know, it takes a much shorter exposure to record them. Under moonlight, good results may be obtained by using objectives at focal ratios of about 1.5, extra-sensitive panchromatic emulsion, and an exposure of about 10 seconds, depending on the circumstances. These times are short enough to avoid effects of movement caused in the subject by winds: cloud movements and turbulence are not picked up. Landscapes may also be photographed by the same techniques, using, according to conditions, exposures of about 15 to 30 seconds at the most.

Study of Figs. 266 and 267 will show just how far the Moon may be used as a substitute for the absent Sun.

THE MOON'S PHYSICAL CONDITION

General appearance of the Moon. — Even with the naked eye, it is possible to make out certain details of the Moon's surface:particularly the large grey areas which remind people of a human face — the so-called 'Man in the Moon' — also differences in intensity in the light areas, as well as irregularities along the edge of the phase (the terminator) which are due to differences in the level of the ground. Thus with the unaided eye we can already begin to detect the diversity of the Moon's structure. The

Fig. 267 — LANDSCAPE PHOTOGRAPHED BY MOONLIGHT WITH AN EXPOSURE OF 30 SECONDS

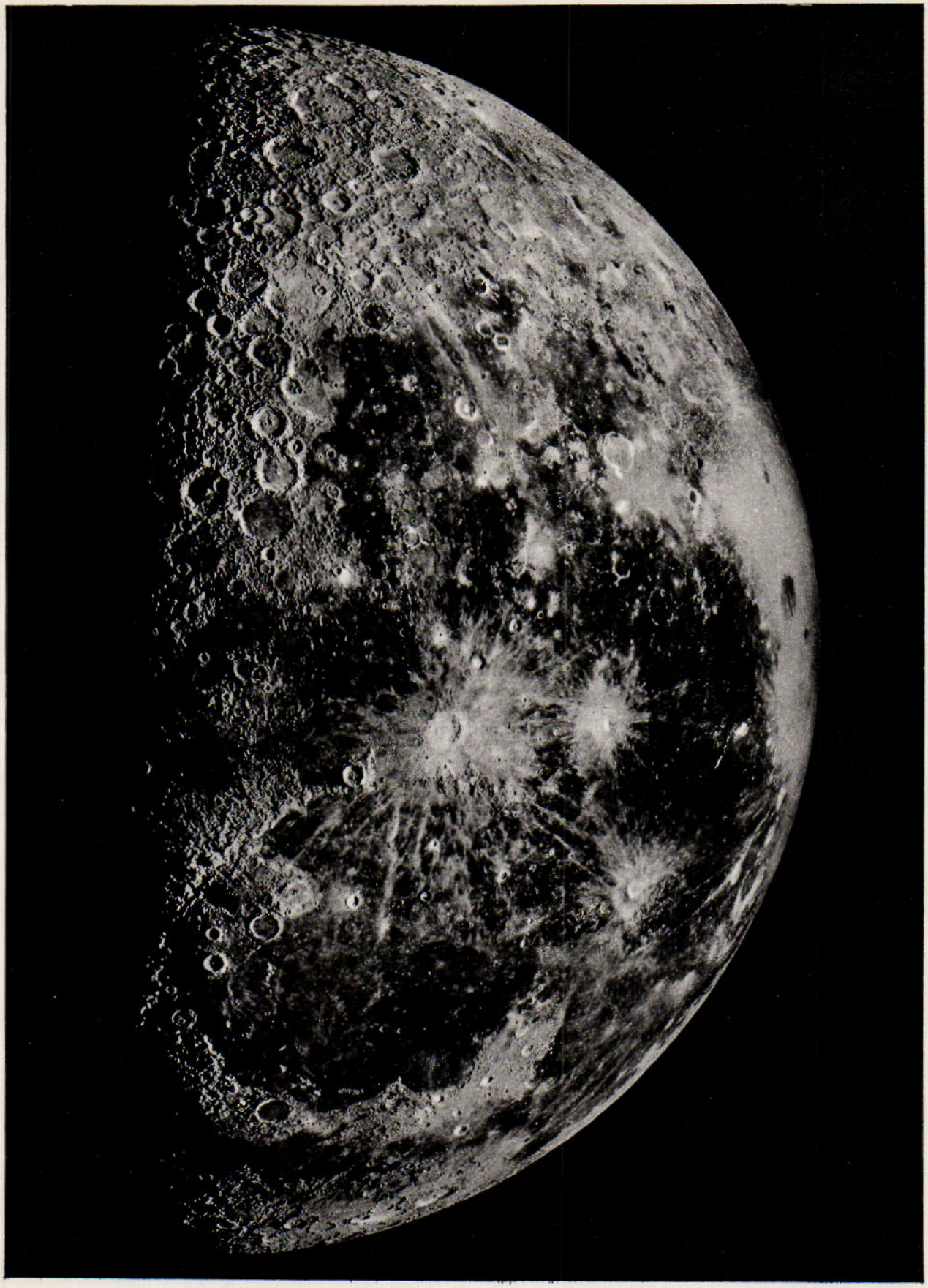


Fig. 268 — THE MOON IN ITS LAST QUARTER *(Photo, Mount Wilson Observatory)*

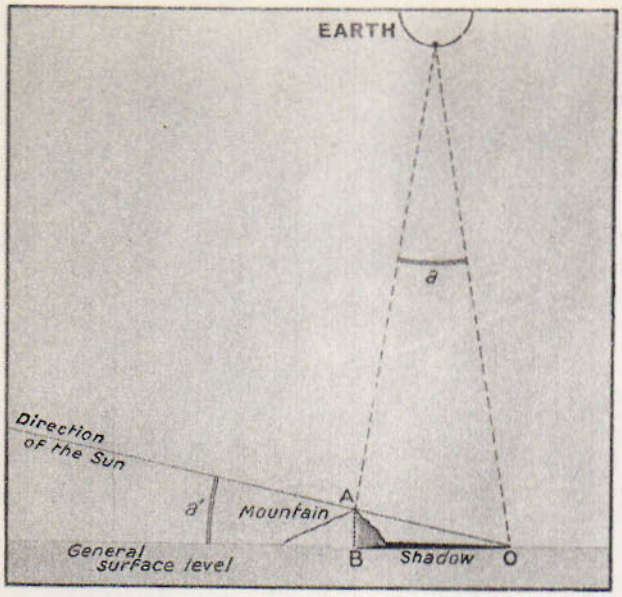


Fig. 269 — THE PRINCIPLE EMPLOYED IN MEASURING THE HEIGHT OF A LUNAR MOUNTAIN

smallest telescope or field-glass, however, is capable of revealing a sight of such beauty that even the most indifferent observer is liable to be seized with excitement.

A powerful telescope will show the rugged surface of the Moon in remarkable detail, and we now have a very minute knowledge of how the various surface structures are arranged, their size, shape and height.

We shall more easily understand the nature of the lunar surface through the medium of photography, which reproduces it exactly, than through words. Our first impression is probably that the Moon seems very crowded with volcanic craters of all sizes and shapes; we are confronted with an absolutely chaotic jumble (or so it seems) of mainly circular formations, packed tightly together in some places, more isolated elsewhere. The scene is made more complicated by numerous mountains jutting up in great chains and splendid massifs. Because of their white colour, these upland areas contrast sharply with the large grey areas, which are vast depressions. Unfortunately there is not sufficient space here to give a complete list and description of all the formations: this would fill a large atlas. Instead we shall summarise the various types of features where they are revealed by telescopic observation, visual or photographic, to good advantage.

The first and most important thing to allow for in lunar observation is the state of illumination, which depends on the phase. As will be remembered from an earlier section, the phase is determined by the Moon's motion, and the surface features undergo considerable changes in illumination following the variation in the angle of the incident light. When the Sun is just rising or setting, its light is coming in almost horizontally, and so every slight irregularity casts a long black shadow, throwing the detail of the landscape into exaggerated relief. On the other hand, when the Sun is overhead, all appearance of relief vanishes, leaving a flat-looking surface with only differences in tone to suggest the presence of craters and mountain ranges. In the former case, the length of the shadow is greatest when the Sun is at its lowest in the lunar sky, and circular formations especially appear as bottomless pits of the deepest black.

Although shadows may sometimes be something of a nuisance (since if too long they hide interesting details from view) they are extremely valuable in that they provide a way of calculating the height of the mountains which cast them.

Formations near the centre of the Moon's disc are seen in plan, as if the observer were hanging suspended at a great height immediately above them. This is the sort of view that an airman or balloonist gets when looking straight down towards the ground. But formations near the edge or limb of the lunar disc are seen more obliquely, in perspective, sometimes almost edge-on. This may be seen clearly in some of the full-disc photographs in the text; at the centre the craters appear circular, but gradually become elliptical as they near the edges, and the nature of the relief begins to stand out more clearly. At the limb itself, the silhouettes of mountains are seen correctly edge-on, just as they would appear on a distant horizon on the Earth. As on Earth, too, the mountains pile up, one behind the other, in such confusion that it is impossible to see the actual surface level of the ground, or, for that matter, the lower slopes of the mountains. In such conditions therefore, and when the arrangement of the landscape is suitable, we can take direct measurements of the apparent height of the relief, whereas in areas that are only seen in plan, heights must be determined by a series of geometrical calculations.

Determination of the heights of lunar mountains. — The principle whereby the heights of lunar mountains may be determined is shown in Fig. 269.

From the Earth, E, we measure the angle *a* subtended by the apparent length BO of the shadow cast by the mountain in question. Since we know the Moon's distance and the angle subtended by the shadow, a simple calculation gives us the actual length of the shadow. Once this is known, a simple calculation is necessary to calculate a′ which is the height of the Sun above the horizon of the point in question. Thus we have a right-angled triangle, of which one side is known (BO, the length of the shadow), this being contained between two known angles: the right angle ABO, and the angle AOB between the direction of the Sun and the ground. These data enable us to calculate the length of AB, which is, of course, the height of the mountain.

The above explanation is theoretical; in practice the matter is complicated by the curvature of the lunar globe and effects of perspective which distort any object not in the centre of the disc.

We must therefore make mathematical allowances for these factors in order to arrive at the correct values. On the other hand, at the centre of the disc it becomes possible to measure altitudes directly. Thus we have a series of altitude measurements over the whole surface of the Moon, more or less, and we have from this almost as good a picture of the lunar relief as we have of that on the Earth. However, an amateur astronomer, using accurate measuring instruments and some fairly easy mathematics, could do some really useful work checking the already known heights and measuring new ones.

One distinction must be made, however. Heights on Earth are related to sea level; in a later section we shall see that the Moon possesses no water and there is therefore no sea level.

The measured height of a lunar mountain top therefore corresponds to its height above the surrounding surface, just as if, for example, we measured the height of Everest not from sea level but in relation to the Tibetan plateau, which is itself several thousand feet high. Thus it is rather difficult to make a direct comparison between the Earth and the Moon in this respect. Nevertheless, because of its smaller diameter, the Moon has the advantage, its surface relief being, in proportion to its

diameter, the higher of the two. Let us now go on to make a more detailed examination of these surface features.

The lunar 'seas'. — The lowest areas on the Moon correspond to those large grey patches visible to the naked eye. Seen thus, or even with a small telescope, they give the impression of being large stretches of water on a map of the world.

And so, simply on this imagined resemblance and perhaps a few preconceived ideas, the first telescopic observers of the Moon drew maps containing such terms as oceans, seas, gulfs, lakes, swamps, to designate the various features that they saw. With their, to us, mediocre telescopes, which gave neither good images nor high magnification, these old observers were not able to recognise the true nature of the Moon's surface. This was understood only after considerable improvement in optical techniques.

Generally speaking, the 'seas' (or *maria* [Latin]; singular, *mare*) are large, arid and roughly circular plains sunk below the general surface level. Apart from a certain number of mountains and craters which dot their surface, they are fairly even. There is no trace of water, as we shall see later. Although our modern lunar maps still retain the old nomenclature, implying the existence of water, this is merely a convenient descriptive terminology and carries no other implications. It may be noted in passing that the lunar nomenclature still shows the influence of ancient beliefs concerning the Moon's supposed effect on fertility, dreams and sleep, etc. We shall continue to use these picturesque names, emphasising that they do not imply the existence of lakes, seas or oceans on our satellite.

The *maria* may be divided roughly into two groups. The first are circular in outline, usually bordered by mountains, and their grey material ends at a fairly well defined limit, rather like an actual sea-coast. Those of the second group are less definite, merging into surrounding areas with no obvious dividing line between. *Mare* material is often found in small patches between mountain massifs and craters. Indeed the flat floors of many craters would seem to be continuations of the *mare* surface, enclosed within a circular ring of mountains. This, however, forms part of the larger problem of how the craters were actually formed, and will be dealt with in a later section.

The fact that the *maria* as a whole can be thought of as immense flat plains is due above all to the contrast between them and the rest of the surface, which consists of an indescribable confusion of craters, mountains and valleys, so densely clustered together in places they frequently interrupt one another.

However, it is not in fact quite valid to compare these desolate expanses of the Moon's surface with a flat sheet resulting from the freezing of a liquid, for example, for they contain considerable differences in level. There are vast depressions, often of unknown extent, since their angle of slope is so gentle. They are most easily seen near the terminator, where the oblique rays of light always throw up into sharp relief any changes in the level of the ground (see photographs). We can then see numerous ridges of varying height and steepness, often running parallel to the edges of a *mare*.

On the other hand, certain features stand out with remarkable distinctness: these are the immense cracks on the surface known as 'rills'. These rills are often angular and sometimes of great length, even as much as several hundred miles in some cases. They are usually of fairly constant width along their length, some being extremely narrow whilst others may be a mile or two across. In

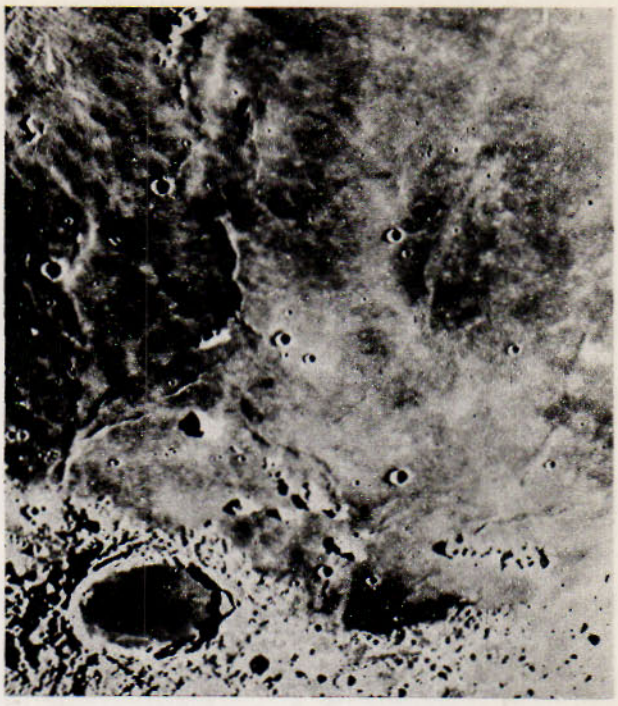


Fig. 270 — THE LUNAR 'SEAS'. Structure of part of the Mare Imbrium.

places there are faults, where, following the occurrence of a crack, the land has subsided or risen to one side, producing something like a step. Although these all seem to be surface features, in many instances they are closely connected with the mountain and crater systems, and also the *maria* nearby.

To many observers, the general nature of the Moon's surface suggests that the *maria* are vast outflows of some fluid material which first partially or completely submerged certain parts of the surface and then solidified. After this it was disturbed by internal pressures and movements of various kinds, some long-term and gradual, some swift and sharp.

Other astronomers have held that these large basins are the floors of ancient, dried-up oceans or seas. We shall discuss this matter further below (page 136).

The lunar mountains. — Any raised portion of the ground may be classed as a mountain, and we can therefore say that the mountainous regions occupy as great an area as the *mare* regions. The mountains of the Moon fall into many categories: for the most part, they form the walls of the characteristic ringlike formations known as 'craters'. These we shall deal with separately. The rest are ordinary mountains, no different in general appearance from those found on Earth. They are sometimes found singly but more often ranged into chains and massifs. These comprise the highest summits on the Moon.

The most majestic of these mountains occur in the region of the south lunar pole, where there are mighty ranges which culminate in the Leibnitz Mountains, about 30,000 feet in altitude, and possibly the highest points on the whole visible surface of the Moon, whilst the Doerfels rise to well over 20,000 feet. The Rook Mountains, which are brought into view by libration on the extreme eastern edge of the Moon, contain summits of 16,000 to 26,000 feet. The height of Mount Huygens in the Apennine chain is 18,500 feet, whilst there are lower ranges of 12,000, 9,000 and 6,000 feet, the latter figure being about the height of many isolated peaks.

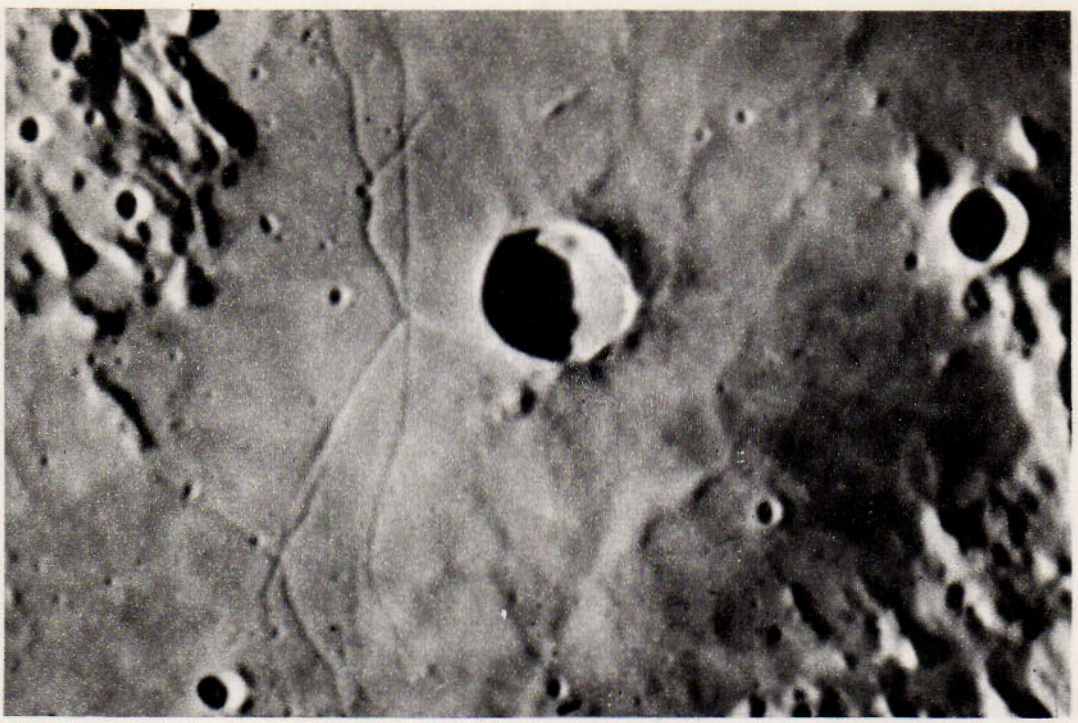


Fig. 271 — LARGE FISSURES IN THE MOON'S SURFACE: THE AREA ROUND TRIESNECKER. Photograph taken by H. Camichel at the Pic du Midi Observatory in 1945. (On this scale the diameter of the entire Moon would be nearly 9 feet).

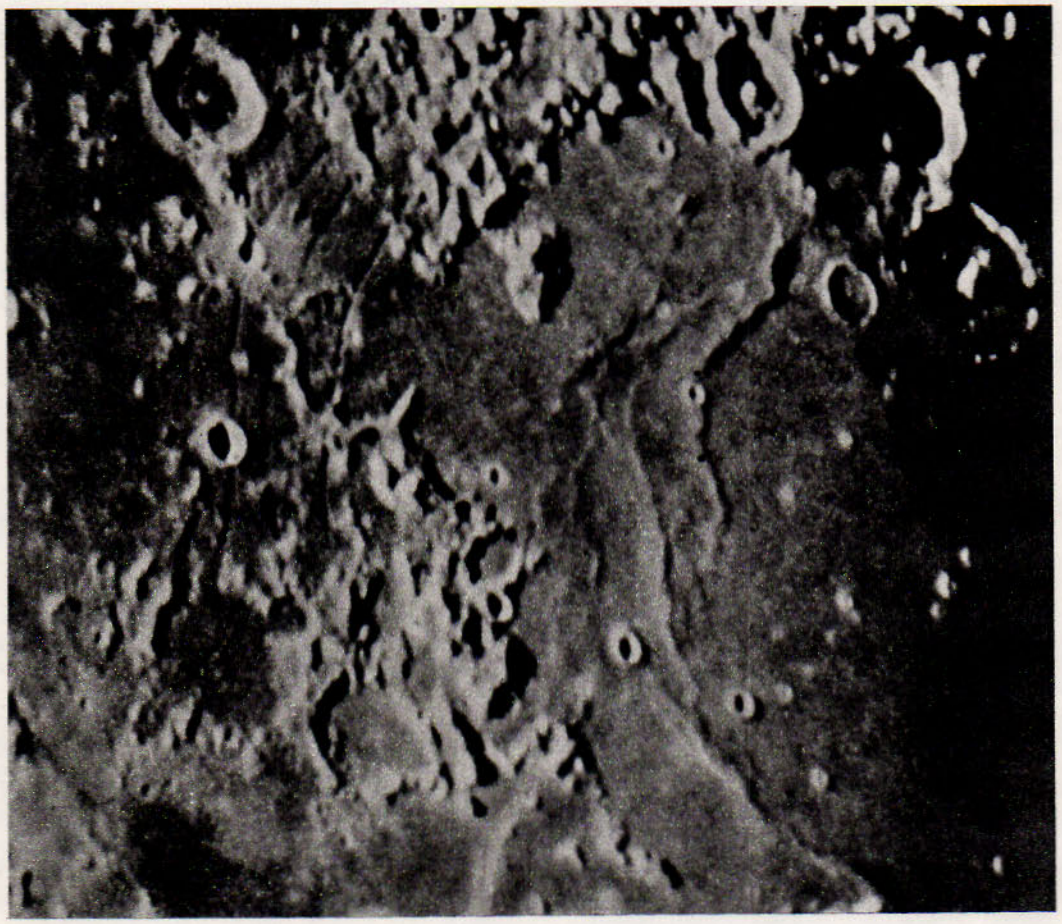


Fig. 272 — GREAT PARALLEL CREVASSES ROUND THE MARE HUMORUM. Note the ridges on the surface of the *mare*. *(Photo, Paris Observatory)*

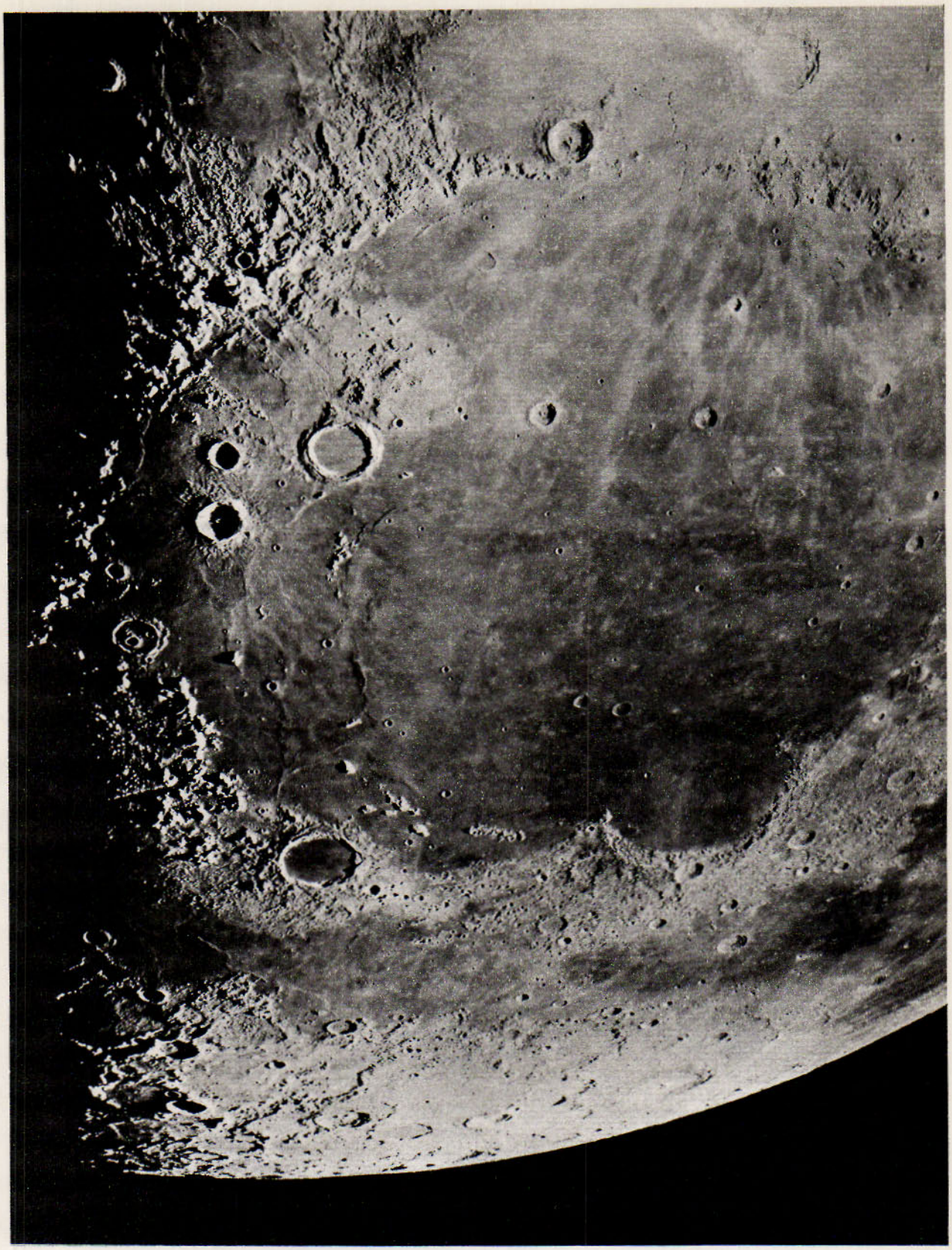


Fig. 273 — CHAINS OF MOUNTAINS ON THE MOON: APENNINES, CAUCASUS, ALPS, ON THE TERMINATOR TO THE LEFT OF THE MARE IMBRIUM. TOP RIGHT, COPERNICUS; BELOW COPERNICUS ARE THE CARPATHIANS.
(Photo, Mount Wilson Observatory; 100-inch telescope.)

Fig. 274 — GEOMETRICAL RECONSTRUCTION OF A LUNAR MOUNTAIN. Mount Pico, with its long pointed shadow, can be seen in the centre of the smaller photograph on the left. On the right it is shown as an observer on the Moon would see it.

Lunar mountains are not as steep as they appear in the telescope. We detect the presence of elevated portions of the surface by the shadows they cast under oblique illumination, but it must be remembered that a shadow gives a distored impression of a mountain's real outline; in the same way if we measured a man's shadow on the footpath at sunset, we should find that it differed considerably in size, shape and proportions from the man himself. Applying this to the Moon, we must not, when observing a series of long, thin, pointed shadows, think that the mountains which cast them are equally long, thin and pointed. As often as not, the mountains in question are neither high nor sharp; indeed they may be quite rounded in outline (although the finest features are jagged). If we reconstruct geometrically the outlines of certain lunar mountains from their observed appearance, we shall find that instead of being steep and jagged, they have quite gentle slopes and their summits are frequently flat or smoothly rounded. This is confirmed by our observations of mountains seen sideways-on standing out from the edge of the disc.

Massifs formed of a great number of mountains usually seem at first sight to be nothing but a chaotic jumble of matter, haphazardly thrown together. On closer examination, however, a certain structure begins to emerge: well defined ranges, as well as patterns of straight cracks or faults, which correspond to other systems elsewhere on the Moon. Moreover, since the mountain massifs have something of a geometrical regularity, their edges follow the same general direction. These edges are also frequently steep, as if adjoining portions of the surface had sunk almost vertically.

The mountain massifs give the impression of being huge areas which have tilted up bodily at one edge more than the other, above the level of the surrounding country. Thus the Apennines form the high border of the Mare Imbrium, a high jagged rampart rising steeply from the plain below. Elsewhere we may observe a sort of step-like formation, matching the shape of the other surface features. In this case the chain-appearance is visible only in the first slope.

The lunar craters. — We have already mentioned these ring formations, which are the first and most striking objects seen on the Moon when viewed through a telescope. We said they looked rather like volcanic craters, but the description may give an erroneous idea as to their method of formation. We should prefer to substitute some other, more natural word, but the term 'craters' has stuck and we shall therefore retain it.

Whilst having certain broad features in common, the lunar craters are by no means all alike. First their dimensions vary considerably, ranging from the almost invisibly small up to enormous ring plains, which would hold two or three English counties or more. Thus the diameter of Clavius, the second largest on the visible hemisphere, is 132 miles, according to latest measurements, whilst Bailly, which is right on the limb and therefore much less prominent, has the enormous diameter of approximately 180 miles. Many others do not fall

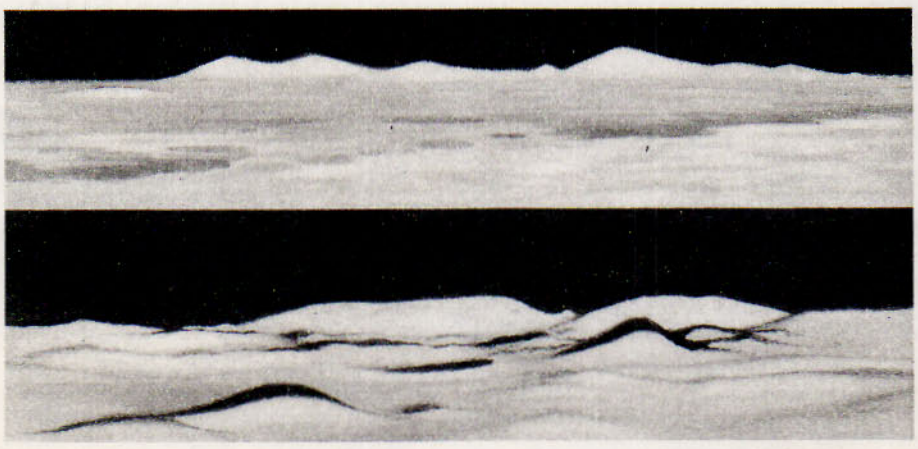


Fig. 275 — MOUNTAINS SEEN IN PROFILE ALONG THE LIMB OF THE MOON

Fig. 276 — THE SOUTHERN REGIONS OF THE MOON ARE COVERED WITH CRATERS OF ALL SIZES. Top right: Tycho with its central peak; bottom right: the great precipice of the Straight Wall. Photograph taken by H. Camichel at the Pic du Midi Observatory in 1944, using the 24-inch refractor.

far short of these two, a diameter of 80 miles being quite common. As the size decreases, the numbers of craters of that size increase in proportion and the number of craters so small as to be on the limits of visibility is prodigious. The smallest visible ones are only a few hundred yards in diameter. The total number of craters with diameters exceeding 1 km. is of the order of 300,000.

Furthermore, the craters are by no means uniform in their structure and circular form. The extremely large formations are usually not so clearly defined as smaller ones; others have gaps in their walls as if they had never completely formed, or as if parts of them had collapsed and disappeared into the still liquid material surrounding them.

There is a further distinction which clearly divides the craters into two kinds: some have flat, uninterrupted floors, whilst the centres of others are occupied by large mountain-complexes. All, however, with certain remarkable exceptions, have one peculiarity which distinguishes them from volcanic craters, with which they have been compared: the floor of a lunar crater is below the level of the surrounding country, whilst volcanic craters on Earth are nearly always orifices opening out from the tops of hills or mountains which have been built up by eruptive action above the general level. As P. Puiseux, who contributed much valuable work on the nature and structure of the Moon's surface, aptly remarked: 'A terrestrial volcano is a raised formation; a lunar crater is a hollow one.'

The heights of crater walls can be measured in the same way as the heights of isolated mountains, i. e. by measuring the length of the shadow they cast. Owing to the sunken nature of a crater, however, the height obtained is greater when measured from the crater floor than from the surrounding country. Taking the craters in which this difference in level is greatest, we find that in the deepest, Newton, the height of the walls above the floor is about 30,000 feet; in the case of Casatus, 22,000 feet; Theophilus, 18,000 feet. Others, especially craters with flat, undisturbed floors, such as Archimedes, are not nearly so high, the maximum elevation being only about 6,000 feet above the interior. We can say that generally speaking the average height of lunar crater walls is somewhere between 9,000 and 18,000 feet above the interior and half of this above the surrounding country. This relationship is not, however, an absolute one, since there are many exceptions, particularly among small formations where the rim hardly projects at all above the general surface level; these give the impression of being mere pits in the ground.

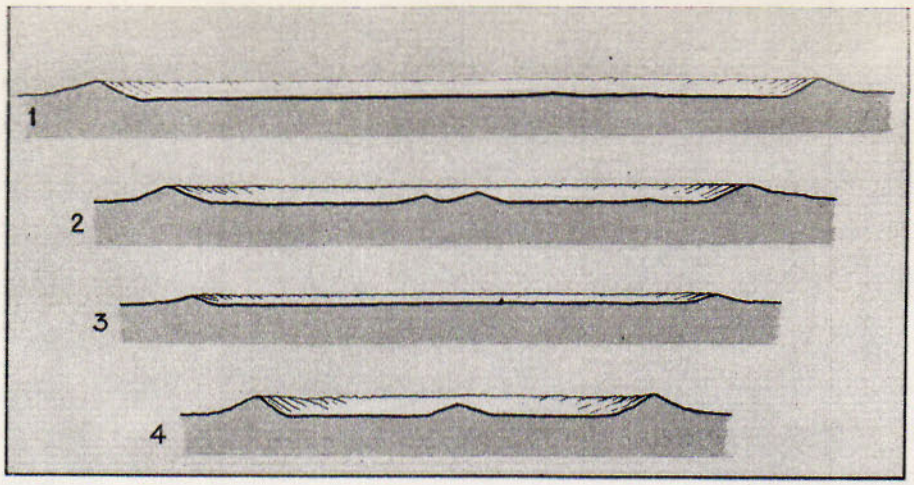


Fig. 277 — SECTIONS OF FOUR LARGE CRATERS. 1, Clavius; 2, Copernicus; 3, Archimedes; 4, Eratosthenes.

We shall have to be content with simply giving some broad indications as to the character of lunar craters, since on account of their immense variety to give a complete description would take up far more space than we have available.

For example, the mean height of the ramparts in certain formations is not the same on both sides. The walls may be interrupted by gaps or by particularly high peaks. And there are considerable numbers of craters which interrupt or partially overlap each other, turning the surface into an indescribable confusion as if the Moon had been the subject of violent convulsions. But this chaos is only apparent. Upon close study, it is possible to pick out relationships both between different sorts of formation and between certain formations and other characteristic surface features, e.g. chains of craters in line with lines of fracture, all forming a sort of grid.

We should not leave this subject without drawing attention to the remarkable systems of white rays which radiate from certain craters and stretch over long distances. For the most part, these rays cross any object in their path without interruption. Moreoever there would

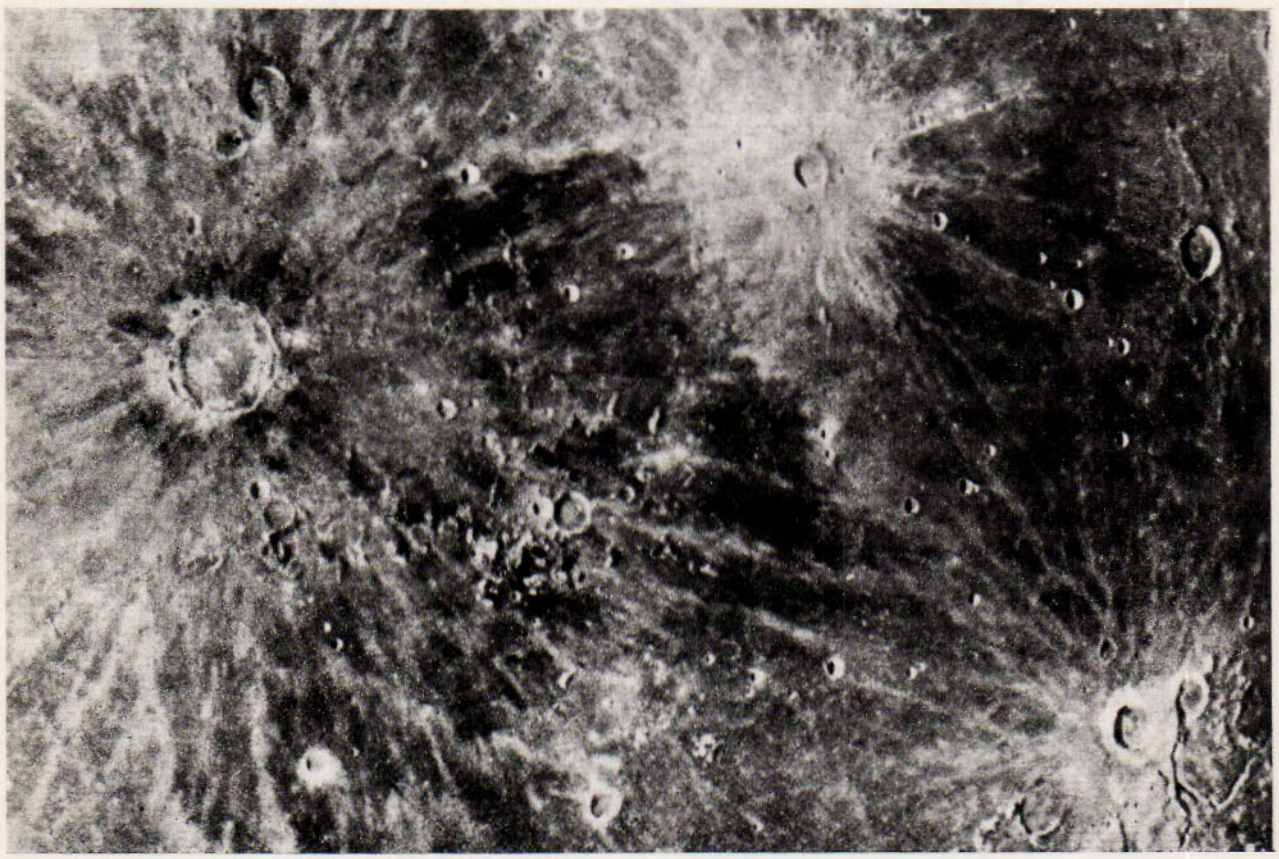


Fig. 278 — WHITISH STREAKS (RAYS) GOING FROM ONE CRATER TO ANOTHER WITH NO APPARENT BREAKS IN CONTINUITY. *(Photo, Paris Observatory)*

Fig. 279 — THE CRATER COPERNICUS (DIAMETER 56 MILES). This crater is the centre of a large system of bright rays, and is surrounded by many different types of crater-formation. To the north of (below) Copernicus are the Carpathian Mountains.
(Photo, Mount Wilson Observatory; 100-inch telescope.)

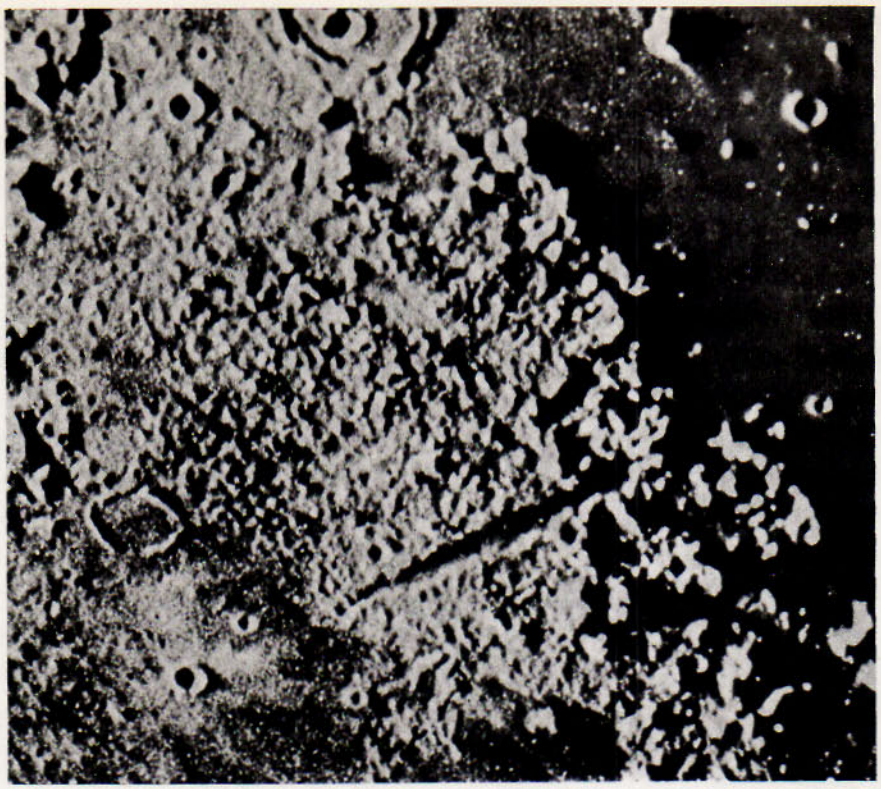


Fig. 280 — THE CHAOTIC APPEARANCE OF THE LUNAR ALPS. Note the absence of valleys such as might have been formed by running water.

appear to be some connection between these white streaks and certain patches on the lunar surface with mosaic-like differences in tone; the rays often seem to form a dividing line between two such patches.

Although it may seem as though we have spent too long in describing these details, we have in fact only sketched in some of the more important features of the Moon's surface. It wil, however, be realised than we can explore our satellite very thoroughly, and such work is highly recommended to the amateur astronomer, who will find it an inexhaustible field of study.

Does water exist on the Moon? — One of the most striking observational facts about the Moon is that whilst even quite minute details stand out quite sharply everywhere, nowhere do we detect the presence of water. In other words, no matter how favourable the observation and illumination conditions, there is nothing which may be attributed to the effects of free water. And although many mountain tops on the Moon are of dazzling brilliance, this cannot, as was once thought, be due to a covering of snow and ice; it is due simply to the nature of the surface.

Instead of open water, then, we might expect to observe the work of rain or rivers, especially as the unceasing action of water is everywhere in evidence on Earth: valleys, gorges, erosion of rocks, drainage of water from high areas to low, deposition of sediment, etc.

On the Moon, however, nothing of the sort is to be found, even when we use a high magnification on a powerful telescope under perfect conditions. No doubt the majority of the features described above would be too small to be seen from Earth, but the grosser features, the large-scale erosion and deposition in particular, would certainly be visible. Now, as we have said, mountain massifs on the Moon take the form of a chaotic jumble of isolated blocks and peaks, and the spaces between them bear no resemblance to the systems of valleys eroded by the natural drainage of water or scooped out by glaciers that we find on Earth. Nowhere, moreover, can we identily a line of mountains which might once have served as a watershed.

There is one objection to all this. The mechanical erosive action of water depends on the speed at which it flows and the nature of the terrain; sluggish water takes a long time to wear away hard rocks. Erosion due to water varies considerably in different parts of the Earth and sometimes its effects are non-existent. This is what might have happened on the Moon, it may be said, for on account of the Moon's low gravity, water would flow very slowly. Moreover, we have very little information about what minerals occur on the surface of the Moon.

In spite of everything, however, the fact remains that water cannot be detected on the Moon, although it would be rash to say that it never existed. The absence of any visible marks only means that its effects must have been slight, and does not disprove its existence at some time in the past. It is just possible that very small amounts of water were once present, but that situation could not have lasted very long. This would explain why the lunar surface gives the impression of having remained intact and unchanged for millions and millions of years.

There is, however, another argument.

When the *maria* were recognised for what they are, it was thought that they might be the beds of ancient dried-up oceans. This hypothesis has a certain plausibility, but we must see if some of the typical lunar features support it or not. What are these features? First, though, before examining the Moon, let us turn to what information the Earth can offer.

If the oceans of the Earth dried up, we should see that their beds are fairly well accidented, as we know from

hydrographic surveys. They would also be found to be covered with certain kinds of sedimentary deposits: alluvial material, sediments deposited and sorted by rivers detritus from marine erosion, sands, muds and oozes of organic origin, to mention only the most important.

These materials tend ultimately to fill in the ocean basins by accumulation, burying and softening slopes and contours generally. Moreover, being of non-uniform composition, they are of a variety of shades and tones, and their various modes and places of origin give them a quite definite geographical distribution. Deposits originating from the land tend to follow the outlines of the continents, broadly speaking, forming a belt all around of varying width.

Let us now return to the Moon. In fact, the observed appearance of the surface is not that which would result from the drying up of oceans under conditions similar to those on Earth, although the presence of certain features such as gentle ridges and swellings would seem to argue in favour of such an interpretation. The immense clefts or crevices would be no objection either, providing we assume that they were formed by internal forces after the drying up of the oceans and the solidification of the sedimentary deposits. However, as we shall see later, not only are the faults and clefts closely connected with the mountain and crater systems, but they would also appear to be intimately connected with the process of their formation. This would lead to the following paradoxical hypothesis: the sedimentary deposits formed from the erosion of the mountain areas were laid down before the latter had come into existence!

We must therefore examine the actual nature of the *mare* surfaces more closely. At first sight they seem to be of a uniform grey tint, but on closer examination by an experienced eye under good seeing and illumination conditions, they show a whole series of slightly different shades, ranging from a faint blue to a vaguely yellowish tinge. It is useless to try to connect these with different sorts of sediments, as on Earth: they do not match up with either the general shape or the relief of the surface features. If anything, they form a faint, irregular chequered pattern, with patches of various sizes, often polygonal in shape. They are often bounded by clefts and whitish rays. It would seem that these too have some connection with the origin of the surface features, and have been caused by the same agency.

All this gives added confirmation that nowhere on the Moon are there to be found indisputable traces of water action, nor does it give any ground for supposing that water has ever existed on our satellite in any quantity.

If water has always been absent from the Moon, the reason for the fact escapes us so far. If water was once present but dried up fairly rapidly, then the reason must be sought in the nature of the ground, or more precisely of the surface layers. We have no information as to the mineral composition of these layers, but the total mass of the Moon is known, and we also know that the mean density of the materials of which it is composed (3.33) is lower than that of the Earth (5.52). Continuing these comparisons, it is logical to postulate that, just as on the Earth, where the core is relatively dense, the surface layers of the Moon are made up of lighter substances with a density not exceeding 2, which is comparable with that of chalk. However this may be, such rocks, which are much less resistant than hard, compact rocks, would be much more permeable to water trying to reach its own level, as on Earth. Moreover, the large number of clefts and cracks to be found on the lunar surface would contribute to such action. Lastly we may draw an analogy between the possible histories of the Moon and Earth with reference to the loss of internal heat. The Moon, being much smaller than the Earth, would have lost all its internal heat much sooner than the latter. All these factors taken together might mean that any water which originally existed on the Moon found its way beneath the surface through cracks and fissures, and since all the internal heat had been given off, and could not therefore cause evaporation, the water remained there.

Has the Moon an atmosphere? — This question is usually answered in the negative, and one frequently hears it stated definitely and simply that the Moon has no atmosphere.

It is worth while to examine this question a little more carefully. Let us see on exactly what grounds such statements are made, since we may find that we have some reservation to make to this absolute opinion.

A first indication may be had by examining the Moon's surface. The tiniest details are always visible with perfect clarity; no definite sign of atmospheric disturbance such as mist or cloud is ever seen, even on a small scale, nor is the phenomenon of twilight, resulting from atmospheric diffusion, ever seen along the terminator.

This immediately shows that if the Moon has any atmosphere at all, it is not enough to cause any visible distortion or veiling of the lunar surface, as would the Earth's if seen from space.

The absence of atmosphere around our satellite can be demonstrated in another way. During the course of its journey round the sky, the Moon will at certain times pass in front of a planet or star (this is called an occultation of the body in question). At this time the body can be seen to disappear behind the Moon's edge, or limb, in the way that it would sink below the horizon on Earth. On the Earth, light from a setting star has to pass through the thickest layers of the atmosphere and therefore suffers maximum absorption and distortion. At the edge of the Moon nothing like this takes place. The limb cuts the Sun at an eclipse, for example, with perfect sharpness, and there are no signs whatsoever of dimming or distortion which might be attributed to atmospheric absorption and refraction. The same thing applies to planetary occultations, and since planetary discs are small, they would probably show such effects more readily. Lastly, to all these negative observations we may add the fact that the spectroscope reveals not the slightest trace of gas or vapour of any sort on the Moon.

It must be admitted that under certain circumstances, some peculiar facts have been reported. These include the truncation and 'hooking' of the crescent Moon, and the presence of a dark band along the edge of the lunar disc as it passes in front of a planet. The former may be explained by the irregularity of the Moon's limb and disturbances in the Earth's atmosphere. The dark band may be the result of an optical illusion caused by contrast between unequally bright surfaces. The same thing is seen when the planet is occulted by the unilluminated limb of the Moon.

Must we then conclude from this that the Moon has no atmosphere at all? No: the most we can safely say is that there is insufficient atmosphere to give rise to any observable effects or to imprint its spectrum on the reflected light of the Sun. This does not prove that no atmosphere whatsoever exists; only that if it does exist, its effects are undetectable and therefore insignificant. Stellar occultations may provide us with some more definite information on this point.

Because of their immense distance, stars possess no measurable diameter. When occulted by the edge of the Moon they disappear instantaneously, and they reappear on the other side with equal suddenness. It would therefore seem to be possible to observe the exact time of

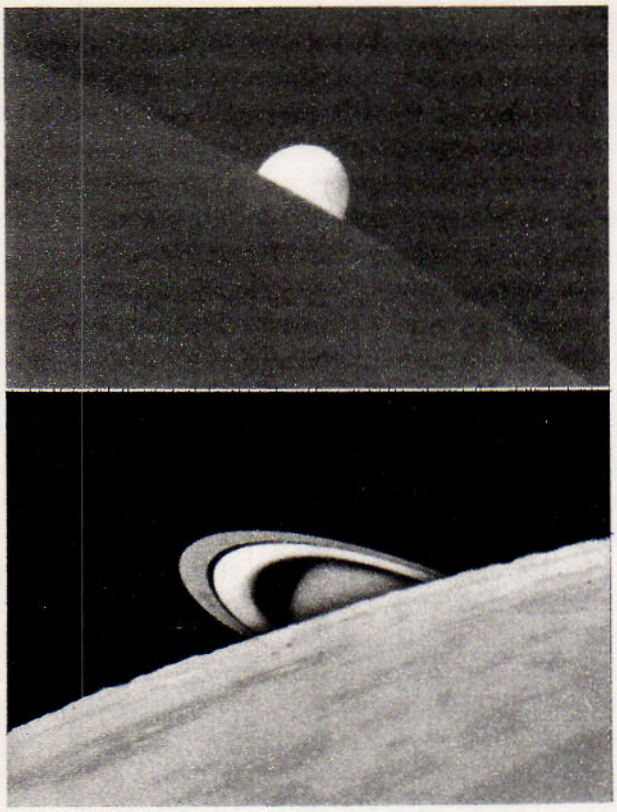


Fig. 281 – OCCULTATION OF PLANETS BY THE MOON. Above: Venus; below, Saturn. The sharpness of the Moon's limb shows that our satellite has no appreciable atmosphere. The dark band crossing Saturn is a contrast effect between two unequally bright surfaces.

disappearance and compare it with the theoretical one calculated from the Moon's known motions. The effect of refraction in the Moon's atmosphere would be to introduce discrepancies between these times; in other words it would alter the Moon's apparent diameter by a very small amount. Two methods have been tried to detect this effect. Based on the directly observed lunar diameter, it would be found that the occultation lasted slightly longer than it ought; on the other hand, when calculated from the duration of the occultation, the diameter should appear to be slightly smaller than when measured directly.

Although simple in theory, this method is extremely difficult to apply, since the effects to be measured are not much more than the observational margin of error. Moreover, the Moon is not a geometrically smooth globe, but a very rough spheriod, and the minute irregularities on the limb are always changing on account of libration. Furthermore, there is the fact that the Moon is a very bright object seen against a dark background, and owing to the phenomenon of irradiation, it appears larger than it should. A number of other, more technical factors must be added to the above, but we shall not go into them here.

The conclusions drawn from this are as follows: in order to overcome the difficulties, the mean values of a large number of occultations must be used, and these seem to indicate a slight difference between the diameters determined by the two methods. This difference tends towards the direction predicted by theory, and would therefore confirm the existence of an atmosphere. The question is: how important is this atmosphere?

From the above differences it has been calculated that its density cannot exceed 1/750 that of the Earth's. Other approaches have given values of 1/2000 and 1/5000 only. The first figure corresponds to a pressure at the lunar surface of about 1 mm of mercury. Because of the low gravity, moreover, it could be expected that this atmosphere would extend to a relatively greater height than that of the Earth.

More recent work using polarimetric methods has shown that the above values are far too high. The greatest possible density would be something of the order of 1/10,000 of the Earth's, but the real density may well be still lower than this. Although it is still possible to maintain that the Moon may have an atmosphere, this atmosphere is for all practical purposes non-existent.

It is interesting to speculate whether this has always been so, or whether it is the result of a slow leakage of air into space. The latter view is in fact the one more generally held, and is said to be supported by the appearance of certain surface features which might be volcanic deposits scattered by air currents; this would mean that at some time in the past the atmosphere must have had quite a high density, but has now leaked away into space or has been absorbed into the interior of the Moon. It may have been reabsorbed following the cooling of the lunar globe, or by all the gaseous and liquid matter having been fixed chemically in the ground. The leakage would have occurred because under the prevailing conditions the mean speed of the molecules would have been greater than the lunar escape velocity (owing to the low gravitational pull of our satellite), thus enabling molecules to escape into space.

Whatever may have been the situation in the past, it is safe to say that at present the Moon is virtually devoid of air and water, and is therefore unfit for the existence of any life as we know it.

Theories of lunar crater formation. – We can hardly examine the surface of the Moon without wondering how all the craters and other surface features came

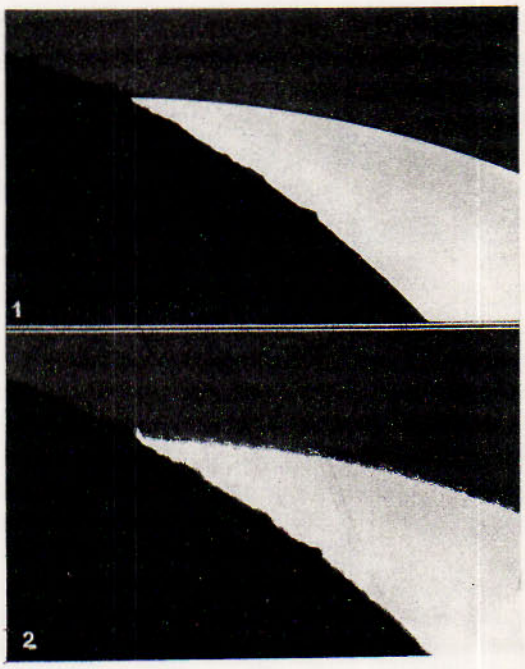


Fig. 282 – DURING AN ECLIPSE THE TIP OF THE CRESCENT FORMED BY THE PARTLY ECLIPSED SUN IS DEFORMED BY AN IRREGULARITY ON THE MOON'S LIMB (1). In the second picture it is deformed by the unsteadiness of the Earth's atmosphere.

into existence. This question also leads us to ask whether any change is still possible or whether the Moon is truly a dead world.

All these questions have at one time or another been subjects for discussion, but none has provoked such heated controversy as the problem of the origin of lunar craters. We shall deal with this first.

The first idea, that they are simply volcanic craters has had to be abandoned, natural as it may at first seem. There is really very little resemblance between a terrestrial volcano and a lunar crater, which is a large circular formation with a very remarkable structure. If, however, we broaden our definition of a volcano to include any outburst of internal forces, we may find more justification. This has the virtue of not forcing us to try to explain every type of formation by the same process. It may well be that the present surface of the Moon has been moulded by swellings and subsidences and eruptions proper (such as we see on the Earth), acting simultaneously or in succession.

One possible mechanism has been suggested by Loewy and Puiseux (the authors of the fine lunar atlas produced by the Observatory of Paris). They envisage a process of swelling followed by subsidence, as shown schematically in Fig. 283.

At a certain stage in the Moon's history, it is suggested that the crust was in such a condition that the expansion pressure of internal gases produced huge, dome-like swellings on the surface. These 'domes' were not stable or permanent, and fissures soon began to appear, whilst at their highest and weakest point volcanic eruptions took place. This released the internal pressure and the 'dome' then began to sink under gravity. This subsidence, taking place in concentric zones, would only leave the outer edge of the 'dome' standing, and this edge would form the rampart of the crater. The terraces which are found in many cases were produced along the edges of the last zones to sink, the main collapse having taken place towards the centre, now either occupied by a mountain mass, or else empty, according to circumstances. Further activity could then take place, giving birth to more craters on the original site or perhaps causing volcanic eruptions on the terrestrial pattern. These would occur principally in the vicinity of the central mountain and along cracks, faults and crustal weaknesses generally.

All this involves the liberation of large quantities of energy, which would result in the fracturing and displacement of large crustal areas. This is confirmed by details to which we have already drawn attention: enormous clefts, furrows, faults and sudden drops, chains of mountains parallel to the mountain massifs. Similar relationships are visible in certain craters, which are clearly polygonal, and in many cases hexagonal in outline.

One thing emerges from this: the most dissimilar features may be linked by a common origin. Some may have been formed before others, or perhaps they occurred simultaneously, on account of previously existing conditions or controlling forces. Many features give extra support to this view. However, as we shall see in a moment, these observations may be interpreted otherwise, leading to quite different views of lunar crater formation but still depending on the action of internal forces.

It should be noted, however, that this fundamental mechanism has not been applied absolutely systematically, since it seeks to explain the origin of the *maria* on quite other grounds.

For example, one theory postulates tidal action, which would have the effect of weakening and dislocating the Moon's crust, thus releasing molten internal material through the orifices so formed. Crater walls and other relief are supposed to have been built up little by little over a long period. Then a general retreat of the molten material allowed the interior to solidify, so producing flat floors. A final outburst gave rise to the central mountain.

Finally, another hypothesis seeks to explain the lunar craters by causes external to the Moon; according to the meteoric hypothesis, the lunar surface has been bombarded by fragments of cosmic debris of all sizes. These fell at enormous speeds, and upon impact their kinetic energy was transformed into extremely violent explosions which produced the craters. The supporters of this hypothesis have cited the famous Arizona Meteor Crater, which was undoubtedly caused by the fall of a large meteorite at some indeterminate time in the past. This crater in every way resembles certain of those on the Moon. But is it valid to generalise from this? We have already seen the dangers of drawing too close an analogy between terrestrial and lunar volcanic craters and, moreover, it is difficult to understand why the Moon alone should have been bombarded so heavily with meteors. Furthermore, the distribution of meteor craters would be essentially random, with no conspicuous arrangement of or relationship between craters. And meteors could hardly have caused the many minor features which are obviously associated with the craters but quite different in structure.

More recently this theory has been modified by taking into account the braking effects of a progressively diminishing atmosphere, as well as the erosive action of ancient lunar seas, if any, the present *maria* being considered as the dried-up beds of the latter. Also considered are the secondary effects of meteoric impact, such as the scattering of dust and debris in all directions, producing rays, grooves and myriads of smaller craters. All this results in a much better attempt to explain crater distribution, size, shape, etc., so that we cannot formally reject the meteoric hypothesis.

However, until we obtain proof either way, it seems inherently more probable that the surface of the Moon has been moulded by internal forces, and following the hypothesis of this nature described above (swelling, collapse and crustal movements), another hypothesis has been suggested and appears to be more satisfactory. Moreover it is similar to the first in a number of points concerned with the effects of fissuring; it differs, however, as to the principle involved in the formation of the relief, whether ramparts, craters or simple mountains.

This theory was put forward by the French astronomer G. Delmotte about 45 years ago. It suggested the exudation of viscous magma, again under the pressure of internal forces, under conditions which allowed it to break through to the surface. At the same time Rudaux was independently thinking along the same lines, having been led to postulate such a mechanism from his long observational work on the subject.

As in the case of the others, we can only give a résumé of the theory here, and we shall confine our attention to the most salient points.

In the first place, the idea that the lunar relief was produced through the upsurging of internal material is supported by the fact that in nearly every case mountains and the like rise up suddenly and sharply from an almost flat surface, and are different in brightness from the surrounding surface areas. In other words, mountains and crater walls do not give the impression of being due to the upswelling of the particular part of the lunar surface; they look as if they are due to the eruption of a totally different kind of material, which has spilled out from the interior. It may be noted, though, that the various ridges which cross the *maria* are simply more or less well marked changes of level which have occurred there.

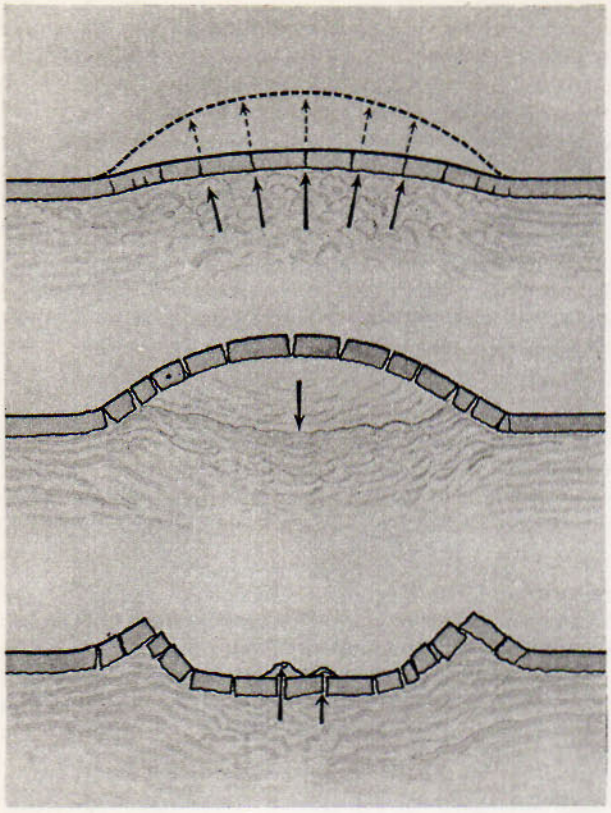


Fig. 283 — LUNAR CRATER FORMATION ACCORDING TO THE SWELLING-COLLAPSE HYPOTHESIS

If the relief were produced in this way, it would mean that the crust must have been in such a condition that the internal magma could penetrate it, which is to say that it must have been fissured over a wide area. Now the evidence is overwhelmingly in support of this view, and it is not necessary to dwell on the matter except to say that there are many features which indicate the heterogeneity of the lunar crust and which would therefore have made it very liable to fissuring. As we have seen, the *maria* have a structure which is broken up into patches or 'compartments', like a mosaic of unequally large but similar pieces. These pieces are roughly polygonal in shape, often with curved sides, and the dividing lines between these areas are often whitish. These joins, which are probably lines of least resistance, might have been the cause of surface or interior fissuring. This would explain the white colour of those rays that are sometimes found on their own but more frequently in radial patterns round certain craters. This would result from their being filled, but not overfilled, with internal magma or perhaps from some chemical action on the ground caused by gases being expelled from the craters in question. Such notions have been current for some time, but relating them to certain quite definite surface features gives them added credibility. It must be mentioned, however, that Loewy and Puiseux saw them merely as streamers of volcanic dust borne on air currents when the Moon still had an atmosphere. This hypothesis requires very violent eruptions and a relatively thick, turbulent atmosphere to carry the emitted dust and débris. Even if we allow for this, however, how do we explain the fact that the appearance they produced has persisted to the present day? If there was sufficient atmosphere to carry the dust, it must have been sufficient to obliterate these surface deposits after a time. Is it possible to believe that the atmosphere suddenly disappeared after the simultaneous formation of all the ray systems, leaving them intact? The tiny white haloes which exist round many of the smaller craters might nevertheless be due to materials thrown out over a relatively large distance, which would be facilitated by the Moon's low gravity. We must remember that similar appearances may result from dissimilar causes, and this applies particularly to the details of the relief that we are about to study.

The fracturing of the crust prepared the way for the underlying magma to be forced upwards by the various internal forces, gradually building up mountainous formations; their size, shape and distribution depended on exactly how and where the magma broke through, i.e. whether the outlets were isolated or members of a system. A circular crater could have been formed out of a series of cracks intersecting in such a way as to form a roughly circular outline. The sunken floor may be explained as follows (see Fig. 284): the eruption of magma to form the walls created a void beneath the floor of the crater, which then collapsed under its own weight, which in turn would probably press out more material through the surrounding cracks. Alternatively these two forces might be acting simultaneously, the one reinforcing the other. If internal pressure detached the floor completely, however, it would then be lifted above the normal surface level and instead of a hollow crater there would result a raised, circular plateau, rather like a table, or more picturesquely, a thin cheese. There are a few cases of such formations on the Moon.

Moreover, many craters are quite clearly polygonal in shape. This no doubt results from the intersection of the cracks which gave them birth. There is a close resemblance between these formations and the shape of the variously toned patches found on the *maria*, marked off by whitish streaks. When the magma was forced out to form the ramparts, it is quite likely that it rounded off the angles, especially if the polygon in question was one with a large number of sides. This would explain both the multiplicity of shapes found on the Moon and the preponderance of round ones. It may be noted in this connection that many present cracks and similar types of fissuring are clearly curved.

The production of terraces on the ramparts of many craters may be explained on the basis of an alternate swelling and collapsing of the interior due to surges in the erupting magma. We have ample evidence of this everywhere, the last examples occurring later than the main moulding of the surface, as indicated by the immense fractures which cut many features or appear in networks and parallel formation with many crater systems. The same thing may also have happened in the interior of craters in process of formation; this would explain the central peaks and other smaller features connected with the varying pressure from beneath. Craters inside or interrupting others may be either later than or contemporary with the primary craters, according to the initial arrangement and size of the fissures. The 'ghost' craters, with partially or completely ruined walls, supposedly submerged beneath the surface of the *maria*, would be explained on this theory by incomplete fissuring of the circumference preventing the passage of material, or perhaps insufficiently fluid magma. This would explain other 'ghost' craters which simply appear as faint rings, or very low ridges, scattered in large numbers over the surface of the *maria*.

Let us next consider the large number of extremely small craters scattered over every part of the Moon. Most selenographers agree that these may be real volcanic craters, in which case they would be different from the large ring-plains. Like real craters, they are found on the

tops of mountains, along the crests of ramparts and where ramparts meet. In addition, they appear in chains along certain crevasses and bright streaks, and at their point of intersection.

If we merely consider the most essential and characteristic features of the lunar relief, the above explanations would seem to be sufficiently convincing. The varied processes they bring into operation are not difficult to accept, since they are due to natural forces found operating in many places on Earth. What we must ask ourselves is if we are not asking too much of them. In an attempt to answer this question, there are a few experiments which can be performed in the laboratory: in other words, by simulating the conditions to which the lunar surface materials were subjected, it may be possible to produce similar effects, naturally on a smaller scale.

This procedure may be objected to on the grounds that whilst it claims to be an attempt to reproduce actual conditions, there is in fact too great a disparity between the forces and materials used in each case. But if our experiments are used only to verify that certain effects deduced from a particular theory can be produced by a given cause, then our procedure becomes defensible. With this in mind, we have here reproduced a number of photographs (Fig. 288) showing certain simple types of formation produced experimentally under conditions attempting to simulate those on the Moon. They were produced by heating water and turning it into steam, the pressure of which was used to force a layer of viscous paste through fissures in a hard surface.

The following conclusions may be drawn: either the fissures simply fill up or else the paste oozes out to form a miniature rampart. When the outline is polygonal, the paste softens and blurs the angles, the intersections giving rise to explosive spurts which form little craters. These minor centres of eruption often occur along fissures like a row of beads.

Once more it may be emphasised that these pictures are intended for comparison purposes and are not meant to reproduce the appearance of lunar craters in detail. If this does sometimes happen (as the reader may have noticed for himself) it is no more than a fortunate coincidence: in a repetition of the experiment, no matter how accurate as far as the general structure is concerned, the appearace may well be completely altered by relatively minor changes in detail.

The fact that we finally return to hypotheses, many of which will have many points in common, should not be surprising, since they are based on the present state of our knowledge, which in turn is derived from visual and photographic methods. As shown earlier, there is still no way of obtaining exact information about the most minute structural details of the craters or the minerals constituting the lunar surface.

Nevertheless, the face of the Moon has been the scene of tremendous natural forces producing most remarkable effects. It may be wondered whether this occurred aeons ago, so that the Moon has remained unchanged ever since, or whether there is still some sort of surface activity which may still produce changes visible through the telescope.

This question has become a matter of some controversy, since certain observations, if confirmed, indicate that changes still take place on the Moon, either periodically or else as single, isolated events. The first question to be decided however, is whether these are due to some sort of physical activity or are purely optical effects. This question is important, and merits being discussed in greater detail.

The nature and physical condition of the Moon's surface. – The exact composition of the rocks forming the Moon's surface is still not known; all that can be said is that the rocks reflect the light very poorly, in a manner similar to basalt. This, however, is too vague and general a statemant to be of much use.

We can infer that there is a great diversity in various parts of the Moon simply from the obvious differences in brightness of the ground. Having nothing else, we must rely on this to guide us.

Generally speaking, the mountainous areas of the Moon are bright in tone, and this contrasts sharply with the sombre aspect of the large flat *maria*. From this we may infer the presence of two sorts of material, with the

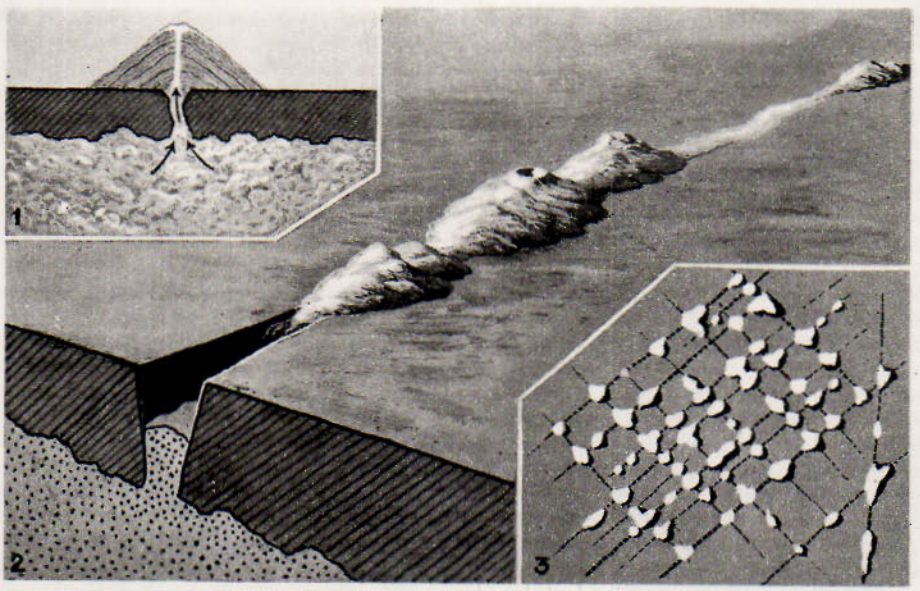


Fig. 284 – FORMATION OF LUNAR CRATERS ACCORDING TO THE HYPOTHESIS THAT THEY WERE FORMED BY A BREAKING-THROUGH OF INTERNAL MAGMA. 1, formation of a mountain (section); 2, relief elongated and irregular, corresponding to the track of a fissure; 3, a fairly regular group of mountains which have burst through a series of intersecting fissures.

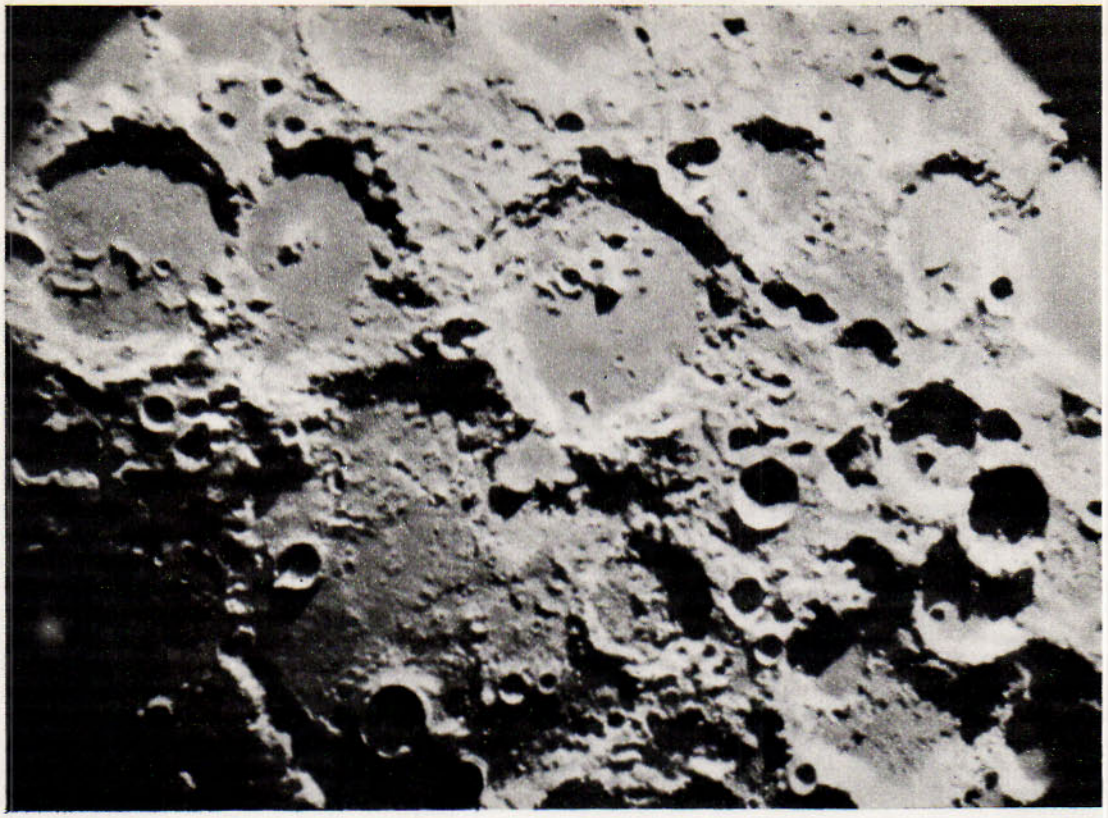


Fig. 285 — A GROUP OF POLYGONAL CRATERS, CONNECTED WITH THE STRIATED APPEARANCE OF THE RELIEF

composition of one being very different from the composition of the material which it has obscured. On closer examination, many more detailed differences begin to appear. Crater walls and mountains generally are usually yellowish, but in many cases they are of a vivid white, sometimes remarkably so. The large expanses of *mare*, which, as we have seen, are divided into a large number of smaller patches or compartments, are very

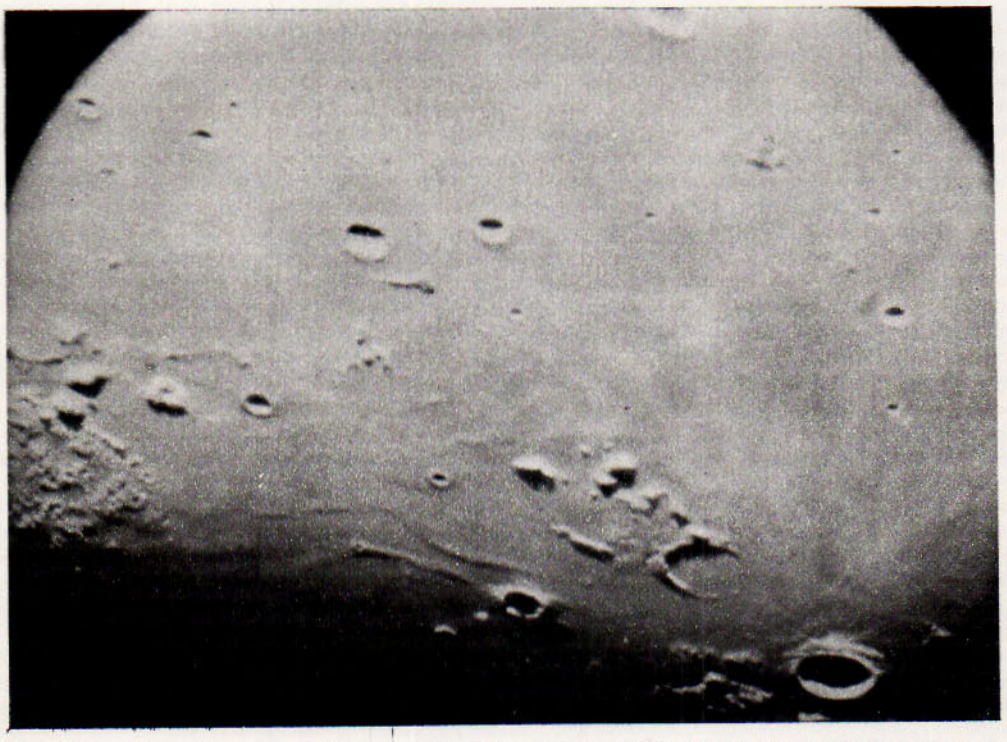


Fig. 286 — 'GHOST' CRATERS. These may be either craters which have never fully developed, or ones which have been buried beneath outflows of lava.

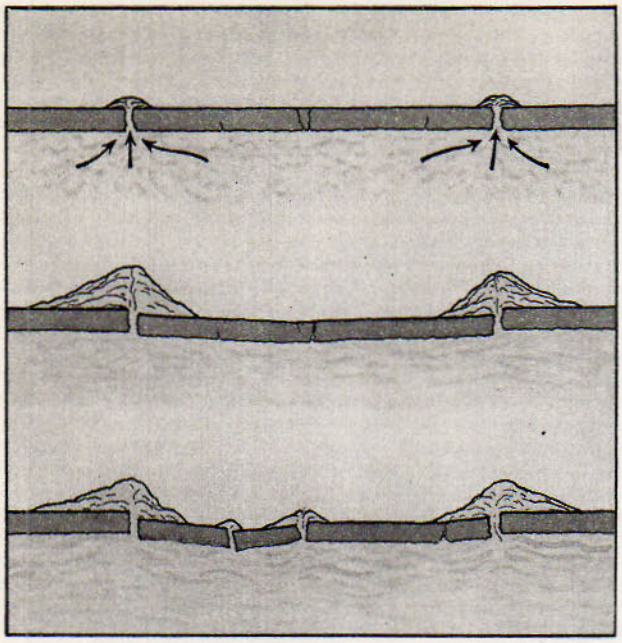


Fig. 287 — HYPOTHESIS OF THE FORMATION OF CRATERS BY BREAKING-TROUGH OF INTERNAL MAGMA

subtly shadowed, with fleeting yellows, greens and slates; elsewhere they are of an indefinable smoke-grey colour, merging into the dark grey of certain crater floors. It may be mentioned in passing that the existence of these patches is an argument against the impact hypothesis: the constant rain of meteoric material ought by now to have dusted over any surface colouring and reduced everything to a monotonous grey — like the furniture of a long-neglected room.

These differences are best seen at full moon because there are then no shadows and all the surface is visible under practically the same illumination conditions, enabling a direct comparison to be made between all the different areas. At this time the contrast between them may increase according to the intensity with which certain areas reflect the sunlight. But this is only a general impression, for the varying direction of the illumination is an important factor. Many features, especially simple patches on the surface, appear to change in extent or in shape according to the direction of illumination. This question is worth examining in a little more detail.

Many factors are involved in the production of these apparent variations. In the first place, the smallest details of the surface relief are not visible in our telescopes; what we may call the 'integrated' appearance of a rough surface will change considerably under different conditions of illumination. When the light is falling obliquely, every little irregularity will cast a shadow, and all the tiny shadows merge, causing a general darkening of the surface; under high illumination the surface brightness will increase. These variations will not be repeated identically because of libration. Furthermore, the angle at which the Sun strikes slopes and hillsides will be different from that at which it strikes level ground. Then again, we have considered only surface irregularities, but the particular reflecting properties of various rocks must also be allowed for: a slate roof, for example, looks much darker under oblique than under high illumination. Something similar may occur on the Moon.

In addition to these varied and variable conditions, there are the effects of seeing the Moon through the thick, dusty, turbulent atmosphere of the Earth: as shown earlier, this more than any other single factor spoils telescopic

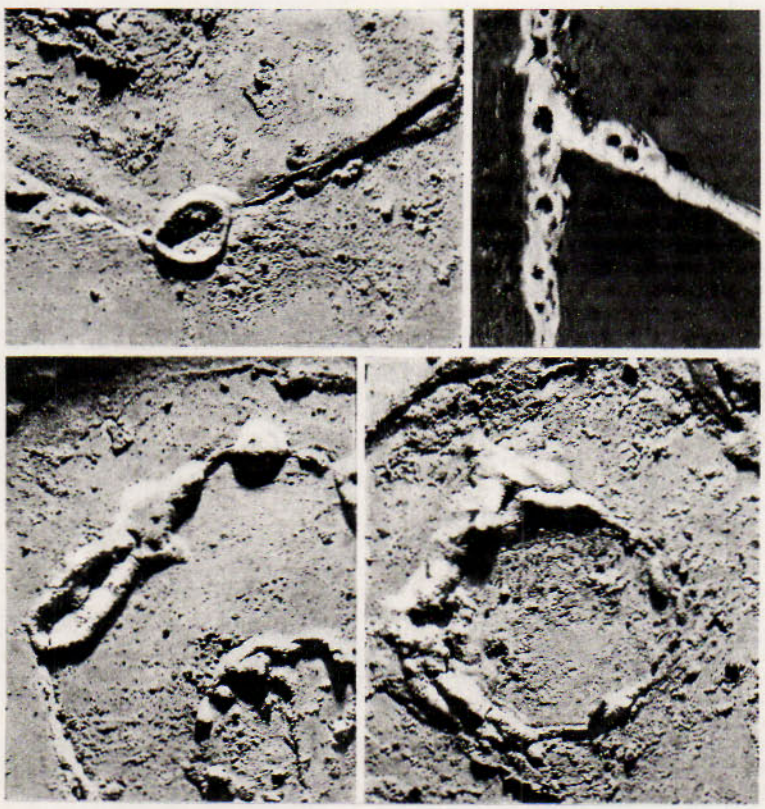


Fig. 288 — CRATERS PRODUCED IN THE LABORATORY BY SIMULATING THE BREAK-THROUGH OF INTERNAL MAGMA

Plate I.

A GENERAL MAP OF THE MOON

giving the names of the principal formations

CRATERS

1. Helmholtz
2. Pontecoulant
3. Hanno
4. Brisbane
5. Oken
6. Philips
7. Neper
8. Plutarch
9. Hahn
10. Berosus
11. Gauss
12. Endymion
13. De la Rue
14. Arnold
15. Borda
16. Euctemon
17. Boussingault
18. Hagecius
19. Biela
20. Furnerius
21. Hase
22. Petavius
23. Nelson
24. Vendelinus
25. Langrenus
26. Webb
27. Apollonius
28. Firmicus
29. Azout
30. Condorcet
31. Cleomedes
32. Burckhardt
33. Geminus
34. Messala
35. Boguslawsky
36. Nearchus
37. Rosenberger
38. Vlacq
39. Jannsen
40. Fabricius
41. Metius
42. Rheita
43. Fraunhofer
44. Stevinus
45. Snellius
46. Reichenbach
47. Santbech
48. Colombo
49. Gogin
50. Gutenberg
51. Messier
52. Taruntius
53. Picard
54. Proclus
55. Macrobius
56. Berzelius
57. Œrsted
58. Atlas
59. Hercules
60. Democritus
61. Kane
62. Peters
63. Ch. Mayer
64. Meton
65. Schömberger
66. Simpelius
67. Manzinus
68. Mutus
69. Baco
70. Pitiscus
71. Neander
72. Piccolomini
73. Fracastorius
74. Beaumont
75. Mädler
76. Capella
77. Isidorus
78. Maskelyne
79. Jansen
80. Vitruvius
81. Le Monnier
82. Posidonius
83. Plana
84. Burg
85. Pentland
86. Cuvier
87. Licetus
88. Maurolycus
89. Stöfler
90. Lindenau
91. Zagut
92. Gemma Frisius
93. Sacrobosco
94. Catharina
95. Cyrillus
96. Theophilus
97. Arago
98. Pliny
99. Poisson
100. Aliacensis
101. Werner
102. Apianus
103. Almanon
104. Abulfeda
105. Delambre
106. Menelaus
107. Linné
108. Alexander
109. Eudoxes
110. Aristoteles
111. Parrot
112. Albategnius
113. Hipparchus
114. Rhaeticus
115. Agrippa
116. Hyginus
117. Manilius
118. Barrow
119. Curtius
120. Maginus
121. Tycho
122. Miller
123. Walter
124. Regiomontanus
125. Purbach
126. Thebit
127. Arzachel
128. Alphonsus
129. Ptolemy
130. Mösting
131. Eratosthenes
132. Archimedes
133. Autolycus
134. Aristillus
135. Cassini
136. Plato
137. Goldschmidt
138. Moretus
139. Clavius
140. Longomontanus
141. Wilhelm
142. Gauricus
143. Wurzelbauer
144. Pitatus
145. Bouillaud
146. Landsberg
147. Reinhold
148. Copernicus
149. Timocharis
150. Lambert
151. J. Herschel
152. Condamine
153. Newton
154. Blancanus
155. Scheiner
156. Schiller
157. Hainzel
158. Capuanus
159. Mercator
160. Campanus
161. Gassendi
162. Encke
163. Kepler
164. Aristarchus
165. Herodotus
166. South
167. Babbage
168. Pythagoras
169. Bailly
170. Phocylides
171. Schickard
172. Doppelmayer
173. Fourier
174. Vieta
175. Mersenius
176. Billy
177. Marius
178. Lagrange
179. Grimaldi
180. Riccioli
181. Hevelius
182. Olbers
183. Seleucus
184. Xenophanes

MARIA

A. Mare Australe
B. Mare Smythii
C. Mare Humboldtianum
D. Mare Foecunditatis
E. Mare Crisium
F. Mare Nectaris
G. Mare Tranquillitatis
G′. Palus Somnii
H. Mare Serenitatis
H′. Lacus Somniorum
I. Mare Vaporum
I′. Sinus Medii
I″. Sinus Aestuum
J. Mare Nubium
K. Mare Imbrium
K′. Sinus Iridum
L. Mare Humorum
M. Oceanus Procellarum
M′. Sinus Roris
N. Mare Frigoris

MOUNTAINS

a. Leibnitz and Doerfels
b. Pyrenees
c. Altai
d. Haemus
e. Apennines
f. Caucasus
g. Alps
h. Carpathians
i. Riphaen Mountains
j. Rook Mountains
k. Cordilleras
l. D'Alembert Mountains
m. Hercynian Mountains

Plate I. GENERAL MAP OF THE MOON drawn by L. Rudaux.
(South is at the top.)

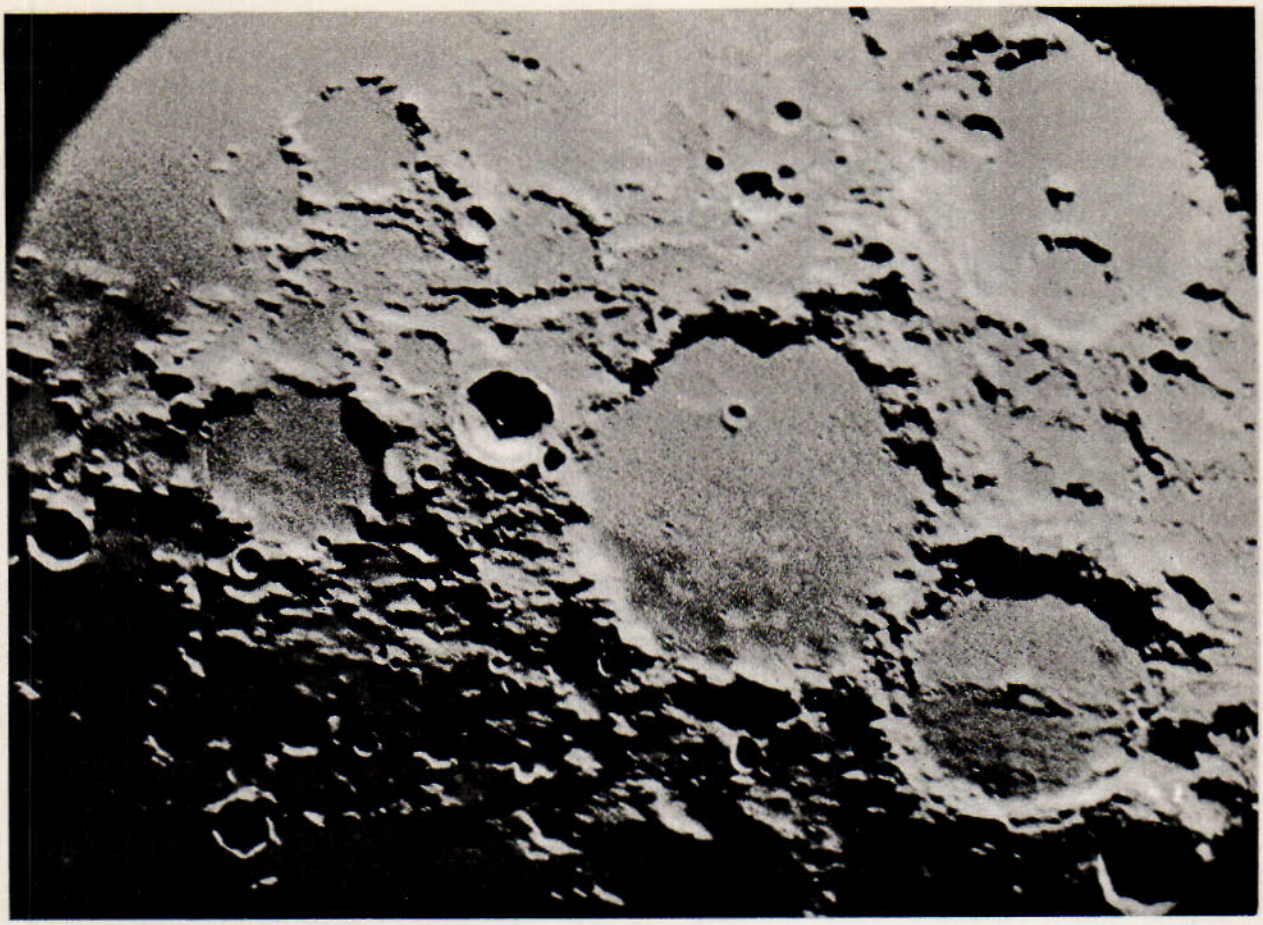


Fig. 289 — IRREGULARITIES ON THE FLAT FLOOR OF A LARGE CRATER

images and misleads us. All these factors, therefore, must be considered before we can say that there is any real, physical variation on the Moon.

The most deceptive factor in lunar observation is the way in which the illumination can completely change the appearance of the landscape. When the Sun is low or at a moderate altitude, a crater will be half filled with shadow and it is perfectly easy to see the nature of the relief. Later, when the Sun is approaching the zenith of the lunar sky, all shadow disappears and the only way of picking out a particular feature is by the surface shading. Thus what a few days before was a prominent ring of mountains will now appear as a series of whitish markings bearing little resemblance to a circle and possibly not even continuous if part of the wall happens to be of the same tint as the crater floor or the exterior surface. Moreover, certain portions of the lunar surface become darker at full moon, and stand out sharply as black patches where nothing of the sort was seen before.

This is probably sufficient to show why people have thought that real changes have taken place simply from a series of well-spaced observations. The most famous example of this is furnished by the twin craters Messier and Pickering (formerly known as Messier and Messier A). Two famous observers, Beer and Mädler, who produced one of the first and best maps of the Moon (1830—1837), show the craters as twins, identical in every respect. Shortly afterwards, in 1842, Gruithuisen noticed that this perfect similarity had given way to a complete dissimilarity of size and shape. This appearance has since been abundantly confirmed photographically, but variations have been noticed which occur as a function of the illumination. The difference in appearance, which is striking under low illumination, tends to disappear under high illumination. Does this in fact mean that Beer and Mädler made a mistake and that no change has taken place? It seems likely. In recent years the behaviour of Messier and Pickering has been intensively re-investigated by Baum and Guest of the British Astronomical Association, who confirm the periodicity of the changes and attribute them solely to optical effects.

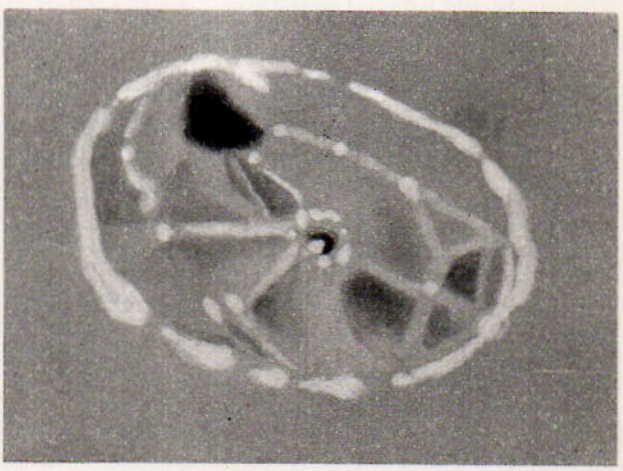


Fig. 290 — ONE OF THE DARKEST PATCHES ON THE MOON: THE INTERIOR OF THE CRATER ATLAS, shown here as seen at full moon.

If such real changes did take place, however, they could be due only to fairly large-scale disturbances such as subsidence, and it would be very surprising if this were not associated with other effects visible to observers from Earth, or occurred so quickly as to escape notice.

Much has often been made of the sudden appearance of surface details where nothing was seen in the past, or of a known object suddenly vanishing. Anyone familiar with the lunar surface, however, is aware of the incredible complexity of detail visible and knows how easy it is to omit or misrepresent a formation when drawing even a limited area. Only a series of good photographs would be acceptable as proof, since the camera is much more reliable than the human eye, but even photographs would have to be spread over a long time and show the disputed area under every possible condition of illumination. Remembering what was said on this matter earlier, we may consider the myriads of extremely tiny crater pits which appear and disappear according to the condition of illumination with no real periodicity. These variations have occasionally been attributed to some sort of volcanic activity, for example the emission of vapour or dust which would hide the formation in question. However, if such a cloud were thick enough to hide a crater from view, even supposing it did not immediately disappear into the vacuum of space surrounding the Moon, it is hard to explain why it should not cast a shadow, and, needless to say, nothing of the sort has ever been seen.

Attempts have been made to explain the variations in the dark and light patches in terms of physical changes on the surface. The first idea postulated the deposition and sublimation of hoar-frost in areas where moisture was present. This moisture was supposed to have come from the interior of the Moon through surface fissures. More recently, chemical reactions have been suggested, which in theory is more likely because it is difficult to admit the presence of water in any form on the Moon.

It would be a very valuable aid to our lunar studies if we had any information about the exact nature of the Moon's surface. The resistance of any particular sort of rock to the natural forces which ultimately lead to its destruction depends on its chemical composition and physical structure. All we know about these at present is that there must be considerable variations from one area to another, as attested by their differences in reflectivity and consequent variability, and that they are probably porous and of low density, which may be deduced from physical measurements. The only destructive forces at work on the Moon are the violent extremes of temperature between the very hot day and the bitterly cold night, since there is no atmospheric covering to temper such extremes. Various methods (including the use of radio-waves) have all indicated that the ground temperature at lunar mid-day near the equator is well over 100° C, and may reach as much as 185° C when the Sun is directly overhead. At night this drops to at least — 50° C and probably much lower. The regular expansion and contraction as night falls, occurring perhaps over thousands of millions of years, must cause a slow cracking and fragmentation of the rocks which will ultimately wear away much of the relief by disintegration. This may well explain a number of apparently blurred and softened outlines in various parts of the Moon (if they were not caused by outflows of lava). These forces, however, work much too slowly for their effects to become visible between one observation and the next; moreover, over extremely long periods it is possible that they could destroy the relief but they could not possibly be responsible for raising it. Thus a projecting formation could only be the result of the upward thrust of internal forces, such as a volcanic eruption, as we understand manifestations of this kind. Unfortunately there is as yet no absolutely indisputable evidence for this on the Moon.

The moulding of the surface could have occurred early in the Moon's history; this must be understood before it is possible to understand its present state.

The existence of volcanoes at some time in the past appears to be confirmed by the remarkable observations of the late Bernard Lyot and his friend Audouin Dollfus in connection with the polarisation of sunlight reflected by the Moon. His observations indicate that the lunar surface is not one of exposed lava but is covered by a thin layer of dust resembling volcanic ash. Now, unfortunately, we have to explain how the pattern of tinted surface patches came into being: what processes could make one area look slightly greenish and another smoke-grey? This is not yet understood.

THE LUNAR WORLD

The respective histories of the Moon and the Earth. — The varied nature of the lunar relief and the causes which we think were responsible for its present state suggest that the history of the Moon has been somewhat different from that of the Earth. In both cases, natural forces have been at work according to the actual nature of the two bodies, their mass and hence their gravitational attraction, which plays an important part in many respects.

To study the history of a world implies the question of its time and place of origin, to which, unfortunately, it must be admitted that no satisfactory answer can be given. We have only theories about the origin of the Solar System. Many suggestions have been made, frequently with no common starting point; although so different, all can be supported with some plausibility, but there is no evidence for us to decide between them. We shall not discuss them fully here, but simply pick out the essential features of the most important theories.

There are two main branches of thought on this question. One suggests that the Earth, Sun, Moon and planets all condensed out of some primeval nebula or gas-cloud; the other envisages a splitting of the Earth itself at some remote epoch in the past when it was still in a molten state. This could have occurred either from internal causes or from some accidental disruption. A third suggestion is that the Moon was captured by the Earth,

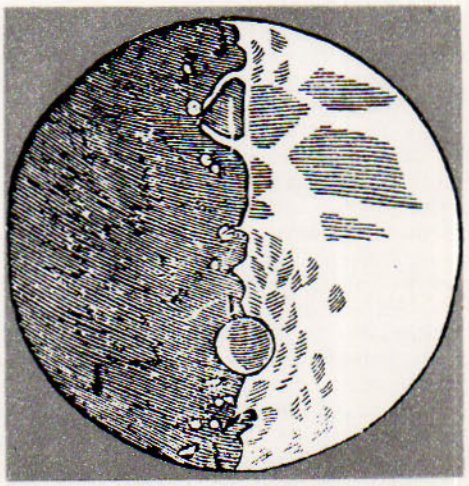


Fig. 291 — ONE OF THE FIRST DRAWINGS EVER MADE OF THE MOON: GALILEO'S (1610)

Fig. 292 — HEVELIUS'S CHART OF THE MOON (1647). The optical instruments in use in the seventeenth century were not good enough to show the real nature of the grey patches visible on the surface of the Moon; they were called 'seas' *(maria)*, and the name has been preserved down to the present day.

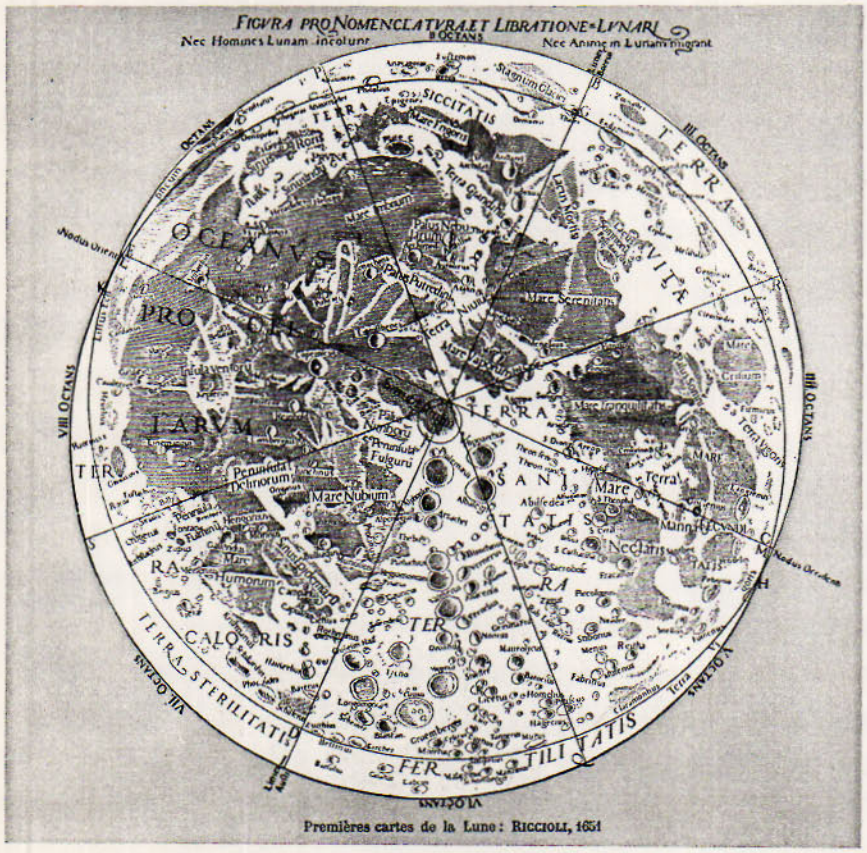


Fig. 293 — RICCIOLI'S CHART OF THE MOON (1651). When the first charts of the Moon were drawn up, every observer used his own system of names, so that the charts of Peirsec, Gassendi, Riccioli and Hevelius were quite different. Riccioli's chart, which was introduced into France by Cassini, forms the basis of the ones in use today. The names of the seas were suggested by the astrological 'influence' of the Moon; the craters were named after great astronomers, philosophers, and so on.

having previously been an independent planet. This however merely puts back the difficulty by one step: the origin of the Moon is still not explained.

What the first two hypotheses really say is that the Moon is either the twin-sister or the daughter of the Earth, the difficulty being that we do not know which.

The exact nature or the relationship being therefore unknown, it is impossible to come to any conclusion about the Moon's origin, although, as will be seen later (Chapter 18), it is difficult to believe that it is older than the Earth.

It will therefore be necessary to begin our investigation at the time when the Moon acquired its present characteristics, bearing in mind the descriptions given earlier of the nature of the surface.

It seems fairly clear (although it is impossible to be definite) that the lunar relief is not due to the same forces that built the mountains on Earth. Furthermore, once the mountains and craters were formed on the Moon, there was nothing to wear them away (excluding the slow and minor effects of exfoliation), whereas on Earth there are very many agents of erosion, the main one being water. On Earth, too, there is a never-ending cycle of uplift, subsidence, ice, running water, erosion and corrosion which continuously changes the nature of the landscape. Everything seems to be evolving relentlessly towards some final state at an unknown but extremely distant date. It is difficult to say just what this final state will be, but it appears that the surface of the Earth is tending ultimately towards a complete levelling off and drying up. Present conditions are transitory, which is really the equivalent of saying that the Earth has a past, present and future. This cannot be said in the same way of the Moon.

The Moon gives every indication of being a fossil world: all development seems to have ended untold millions of years ago. The moulding of its surface must have taken place comparatively rapidly, and once the forces were spent, once the present state of the surface was reached, no further development was possible; none, at least, that might be visible in our telescopes has ever been observed. Thus the Moon may justly be called 'dead'.

It may be wondered whether the Moon is completely dead in every sense of the word. From the physical point of view we can answer: 'Yes'. But the matter may be looked at differently: the fact that no eruptions or natural catastrophes are ever observed does not in itself preclude the possibility that some sort of organic life may be present, even though the conditions found on the Moon would not support life as it is known on Earth. Some astronomers have thought that, in spite of this, certain features of the Moon may be interpreted as being due to the presence of life, and we ought therefore to examine the problem.

Is there life on the Moon? — The question of life on the Moon has intrigued men since the earliest times. Nowadays we can discuss it more authoritatively, since our modern techniques and instruments provide positive information to replace the wishful thinking and speculation of the first astronomers and philosophers.

It would be impracticable to review all the ideas ever published on the subject: some are essentially fantastic, some are more rational. The latter are worth looking at, however, since they occasionally contain inspired guesses at facts which were only arrived at much later by modern science. This does not mean that they described the inhabitants of the Moon themselves but only the surroundings and conditions in which they were supposed to live.

Plutarch, for example, who believed in the existence of 'selenites', thought that rain and wind were lacking on the Moon, the ground alone being able to nourish species of animals different from those on Earth. But apart from this, the famous Greek author's works contain the most improbable collection of ideas.

Anaxagoras went so far as to make the habitability of the Moon an article of philosophical faith. According to him, there were mountains, rivers, lakes and seas just as on the Earth. This was also the opinion of Xenophanes, who thought that the inhabitants of the Moon lived in vast underground dwellings: this opinion was echoed in the nineteenth century when there was believed to be a thin atmosphere around the Moon, the air being most effective where it was compressed at the bottom of deep depressions. Another idea was that the lunar inhabitants lived in great caverns beneath the surface and were thus protected from the rigorous conditions above. Nowadays we know how intensive the fissuring of the Moon's surface must be and what extremes of temperature are met with on the surface. There is, however, one remarkable book which is worth describing: Kepler's *Somnium*.

This astronomer, famous for his three laws of planetary motion, was very familiar with the Moon's movements, and he defined the length of the lunar day and night — one hemisphere, it will be remembered always being turned towards the Earth. He thought that prolonged heating due to the Sun would draw all the liquid round to the daytime hemisphere in the form of vapour and clouds, and he imagined the inhabitants living in vast caves to protect themselves from the alternating intense heat and bitter cold of the surface. Deep canals conveyed hot water through the caverns to refresh them, whilst everything born at the surface grew rapidly and to a tremendous size. Now we have already seen that such extremes of temperature do exist on the Moon, whilst the rapid growth of organisms on the ground has more recently been suggested as an explanation for variations in the appearance of certain areas.

Let us now see how the question stands today.

It is obvious from what has already been said that the Moon is a very different world from ours, in both surface features and physical conditions. Not only are the two fundamental needs of life lacking — water and air — but there is the enormous range of temperature between day and night, shadow and sunlight. Any directly illuminated object in the equatorial regions of the Moon may be at a temperature of more than 100° C, whilst at night this must fall equally far below freezing point. It is well established that highly organised forms of life, whether animal or vegetable, cannot exist outside a very narrow temperature range. There is, however, yet another factor which complicates matters still further.

This is the lack of any atmospheric covering which would protect the Moon from such extremes. In other words, during the lunar day there is nothing to soften or moderate the intensity of the Sun's rays, so that all wavelengths of light reach the ground unchanged. This radiation is very rich in ultraviolet, and when the death-dealing properties of ultraviolet are known the conclusion is inescapable.

Thus with so many hostile factors, life — as we know it — is out of the question on the Moon, and our satellite can be considered one vast, sterile desert.

Some people have thought this conclusion unjustifiably hard. Might there not be other forms of life quite unknown to us, which could exist on the Moon?

This question cannot be decided directly, since we do not possess telescopes powerful enough to reveal living creatures on the Moon, even if such creatures existed. But, we might ask, may they not leave traces of their presence? Might it not be possible to distinguish bare

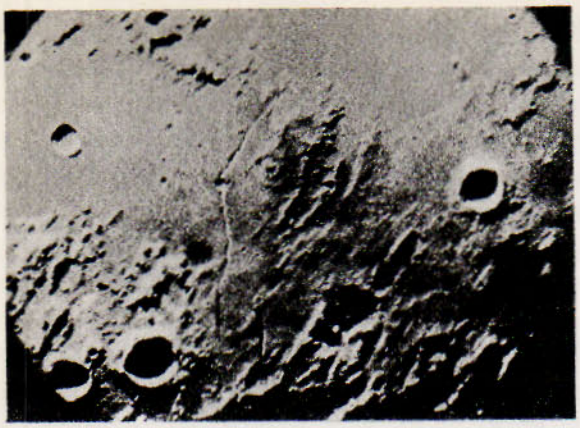


Fig. 294 — Certain features on the Moon's surface, such as white streaks, straight fissures and accidental geometrical figures, were once thought to be the work of intelligent 'Selenites'.

ground from ground occupied by living organisms?

It must be admitted that in the past some details have been interpreted in this way. Such reputable observers as William Herschel, Schräter, Littrow, and Gruithuisen, among others, have thought they could see artificial structures wrought by 'human' hands — fortifications, roads, ditches and canals — which were in all probability the various markings already described: intersecting white streaks, fissures, faults, etc., more or less geometrically regular. But the optical instruments of those days were not so good as ours and it is almost certain that quite natural formations were mistaken for artificial constructions, especially those which show a sometimes astonishing regularity of form.

More recently still and even at the present time, certain areas have, from their appearance, been thought to be covered with some kind of vegetation. This vegetation has been called 'rudimentary' because it was judged more likely to be adaptable to conditions on the Moon than other 'superior' types of organism.

In the first place, attention has been directed towards certain greenish or smoke-grey areas such as the floors of craters like Grimaldi, and towards a number of small, often irregular patches. It has been suggested that this appearance is due to the presence of vegetation, and this hypothesis seemed to be supported by the first photographs the Moon. Because of the nature of the chemicals used, these showed the areas in question with an exaggeratedly dark tint, just like the 'green hills of Earth', which are very unphotogenic, as the amateur photographer well knows. Other arguments, however, have been better founded.

Various astronomers believed that the existence of vegetation on the Moon could be proved by the variations observed in certain markings. This was the opinion of the American astronomer, W. H. Pickering, who studied some curious changes of size and shape as the Sun rose over these areas. If it is accepted that the variations are due to vegetation, this would imply 'plants' which could grow quickly, change their colour and their distribution during the course of a lunar day (which lasts about a fortnight) since the supposed vegetation coverings would vary in tone, and spread over or disappear from enormous areas in a remarkably short time. Moreover, the American astronomer not only maintained that vegetation was present, but also reported observing phenomena which could only take place in an atmospheric medium: clouds, mists, and blizzards, not to mention volcanic eruptions.

It would be very pleasant to believe in all this, but we cannot do so unless it is confirmed with no possibility of mistake. This unfortunately is not the case.

There are numerous observations confirming the fact that variations in tone do occur, but there is no evidence whatsoever for the interpretation that Pickering put upon them. Indeed, his interpretations are open to criticism on many counts, and his opponents were not slow to challenge his findings. As we have seen, the ever-changing direction of the illumination can alter an area almost beyond recognition. This is due simply to the nature of the surface, and it is quite unnecessary to postulate vegetation to account for the observations.

Whatever the truth of the matter, it is a difficult problem to solve definitely, because it depends to a large extent on the delicate, not to say fragile, evidence of telescopic observation.

Perhaps more powerful instruments coupled with more subtle techniques may one day enable us to decide the question one way or the other. In the meantime we must be content with the information we have. This tells us that the Moon is an airless, waterless world, containing no life of any sort, at least at the present time.

However dead it may be, though, desolate and uninviting, one day men will land there and explore the Moon at first hand. Let us therefore try to imagine what it would be like to be a traveller on the Moon.

A traveller on the Moon. — It is possible to make a reasonably good estimate as to the appearance of the lunar landscape, since we can measure the heights of mountains and the sizes of craters and see how the various features are arranged, from Earth. We can then reconstruct their actual appearance as a landscape by the application of the laws of geometry and perspective.

Although it is possible to visualise what the Moon actually looks like for an observer on the surface, it is only the larger features which can be so treated, simply because present instruments are not powerful enough to reveal the smallest details. Nevertheless certain conclusions can be drawn from the observational evidence which may give at least an approximation to the truth.

Moreover, it is not only the shadows cast by mountains and the relief generally which have to be considered; we must attempt to reproduce what novelists would call the 'atmosphere' resulting from the physical conditions existing on the Moon, a setting which any human being would find strange, to say the least of it.

The first thing that strikes us on our lunar journey of exploration is that in the main the mountains are not so steep as on Earth, and their summits are quite often rounded or even flattened. It should not, however, be thought that steep slopes and sudden drops do not exist, for they are found in a number of places where the ground has risen or sunk following the development of a fault or fissure. Mountains on the Moon have one great advantage over terrestrial ones: they are very often isolated and rise straight up from a comparatively flat plain. This sets them off at their best and a lunar mountain must be a truly magnificent sight.

Let us now turn to the craters, about which many people harbour some very erroneous ideas. In the telescope, the Moon's surface often appears to be covered with dark and apparently bottomless pits just as if it had been subjected to a fierce bombardment. The facts, however, are quite different. It must be remembered that by terrestrial standards the craters are very large, even the smallest ones visible in a moderate-sized telescope being only just under a mile across. If we imagine ourselves looking out over one of them, we shall find that it fills almost the entire field of view; if we are outside the crater,

Fig. 295 — THE COMPARATIVE SIZES OF THE EARTH SEEN FROM THE MOON (ABOVE) AND THE MOON SEEN FROM THE EARTH

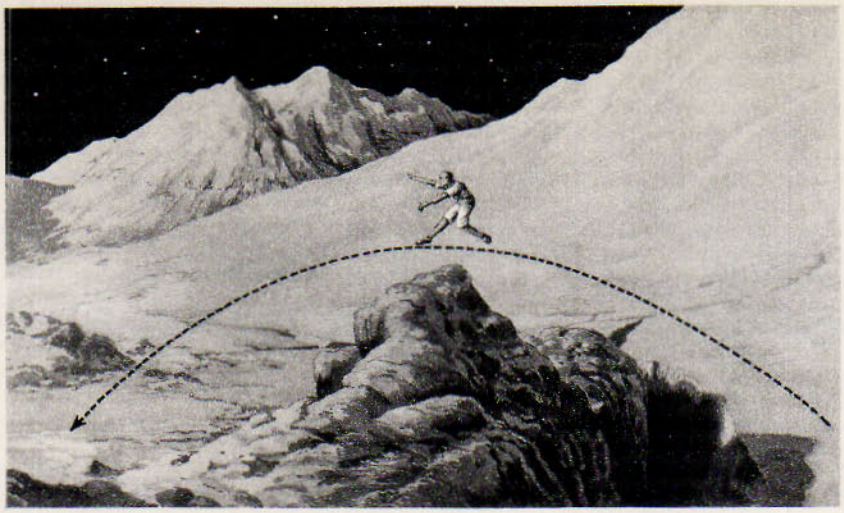


Fig. 296 — FORCE OF GRAVITY ON THE MOON: the jump which an athlete could make, compared with his height.

we will see only the serrated wall of the ramparts, blocking all the rest from sight. To view it as a whole we should have to climb to some vantage point on the terraces, in which case we might also be able to overlook other craters in the vicinity.

When we come to the larger craters, however, this would no longer apply. From the outside it would only be possible to see a long chain of mountains stretching along the horizon. From the inside it would be like being on a vast plain covered with cracks and hills and débris, and the surrounding ramparts would be no more than a line of low mountains girdling the horizon. Thus it would require a positive effort of the imagination to realise that we were inside one of those crater formations which are absolutely unmistakable when seen through the telescope. In the largest formations of all, such as Bailly or Clavius, which have flattish floors, we should be quite oblivious of being inside a crater — unless we had prior knowledge. The flat plain would stretch away to the horizon, far beneath which and invisible would lie the crater walls.

Some of these crater floors, which are frequently so large that they could almost be counted as small *maria*, have an area running into thousands and thousands of square miles. They are terribly bleak and arid, with nothing to break the appalling monotony. Elsewhere, however, the ground is criss-crossed by enormous crevices, which have no counterpart on Earth, at least on such a scale. Although mountains generally on the Moon are not very different from our own in height and general appearance, while crater walls, apart from their circularity, look like mountain ranges on Earth, the clefts are amazingly straight for very long distances, some being several hundred miles in length. With their tremendous width and depth, they would provide both a spectacle and an obstacle of the first order for the lunar traveller.

The above descriptions are concerned with the main features of the very varied lunar surface, but we must now consider the unique characteristic which makes a landscape on the Moon totally unlike anything on Earth: the absence of air. Since there is nothing to diffuse or soften the Sun's rays, the sky is black even in the daytime, and the stars are always visible if the eyes are shaded from the Sun, which shines down with a harsh, raw light in which objects cast jet-black shadows. This stark opposition of light and dark is moderated only by reflected light from the crater walls, which very slightly illuminates the shadows.

Since there is no air, the most distant parts of the landscape stand out as sharply as the foreground. There is no shading off into a misty horizon, as on Earth, and the only way of appreciating the depth of the landscape is to follow one range of mountains as it passes behind another and to note how lines of perspective converge. All this is so different from conditions on Earth that one can well imagine the discomfiture of an orthodox painter setting out to paint such a scene.

The landscape is not the only spectacle, however. The sky is full of strange and beautiful sights, and keeps us fully absorbed. How different the Moon is from the Earth! On our world the atmosphere is responsible for the diffused illumination we call daylight, the beautiful colours of sunrise and sunset, and the mysterious interval of twilight between day and night, not to mention the never-ceasing patterns made by the clouds. All this the Moon lacks. In compensation, the stars shine with an incomparable lustre, and, without the absorption of an atmosphere, the number of faint stars visible is much greater than on Earth, and they seem to fill the sky with luminous dust. The stars never twinkle: that is an atmospheric effect.

The queen among these myriads of lights is the Earth, a splendid disc, thirteen times the area of the Moon as seen by us. During the night of the lunar hemisphere turned in her direction, the full earth is extremely bright (cf. what was said earlier in connection with the Earthlight). The Earth goes through all the phases we see the Moon go through, but in reverse order. The main difference however, is that the Earth remains practically stationary in the sky: from one spot she may always be at the zenith, whilst from somewhere near the edge of the lunar disc, she is always low down near the horizon. Unlike the Moon, whose phases depend on its movement in the sky, the Earth passes through all her phases in the same place. It must, however, be mentioned that to say 'in the same place' is not strictly accurate, for the Earth is not absolutely fixed in the lunar sky; because of libration, which was discussed earlier, an observer on the Moon sees the Earth oscillate for a short distance either side of a mean position. At places where the Earth is normally on the horizon, the librational movement would be easily observed because she would alternately dip out of sight and come back up into view again.

For the same reasons that we only see one side of the Moon, the Earth is permanently invisible from the side we never see.

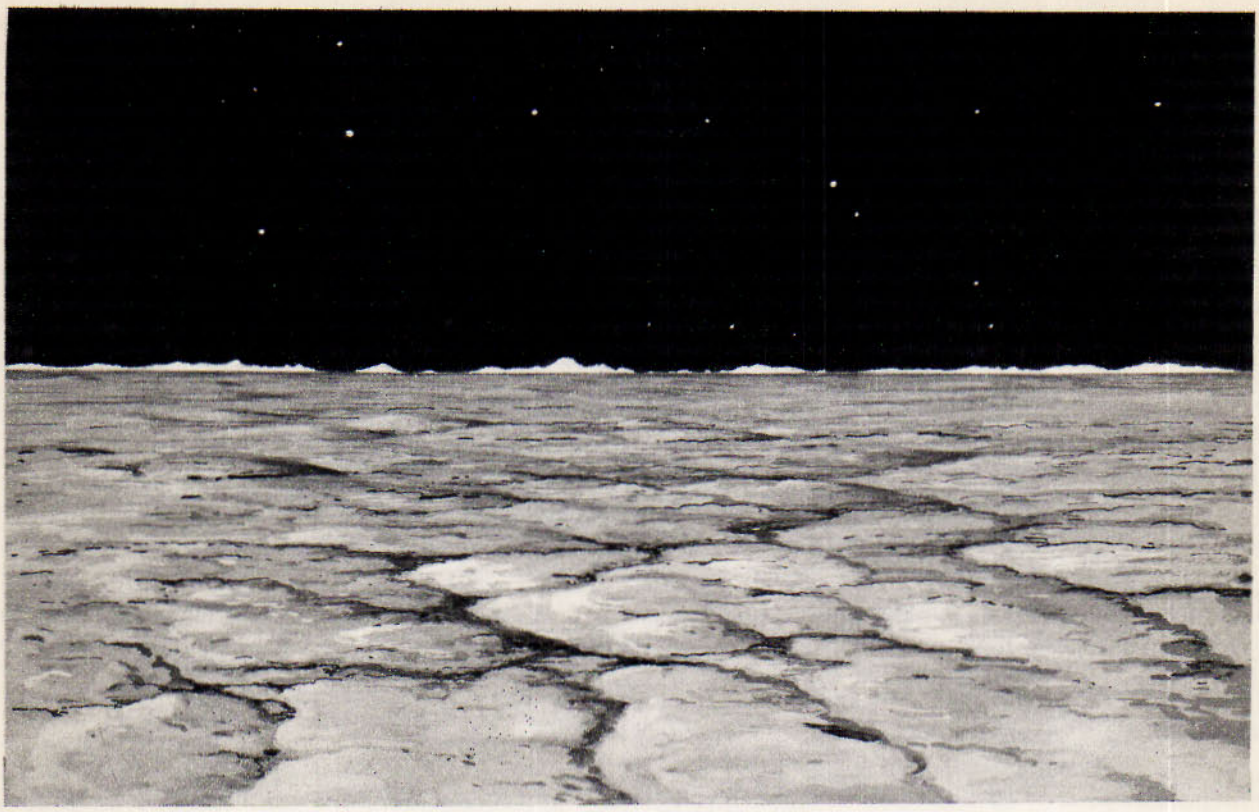


Fig. 297 — A LANDSCAPE ON THE MOON, SHOWING THE APPEARANCE OF A LARGE CRATER WITH A FLAT FLOOR

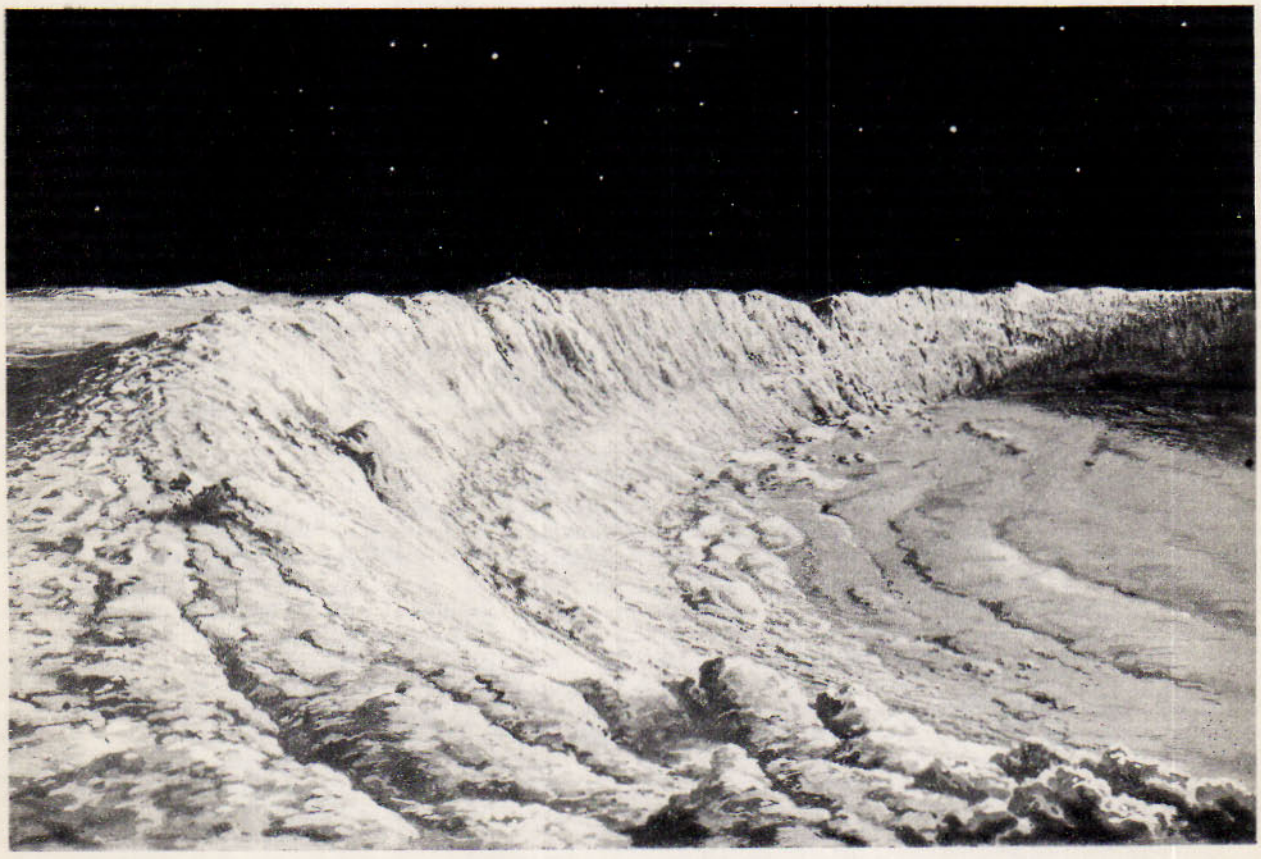


Fig. 298 — GENERAL APPEARANCE OF THE INSIDE OF A SMALL CRATER

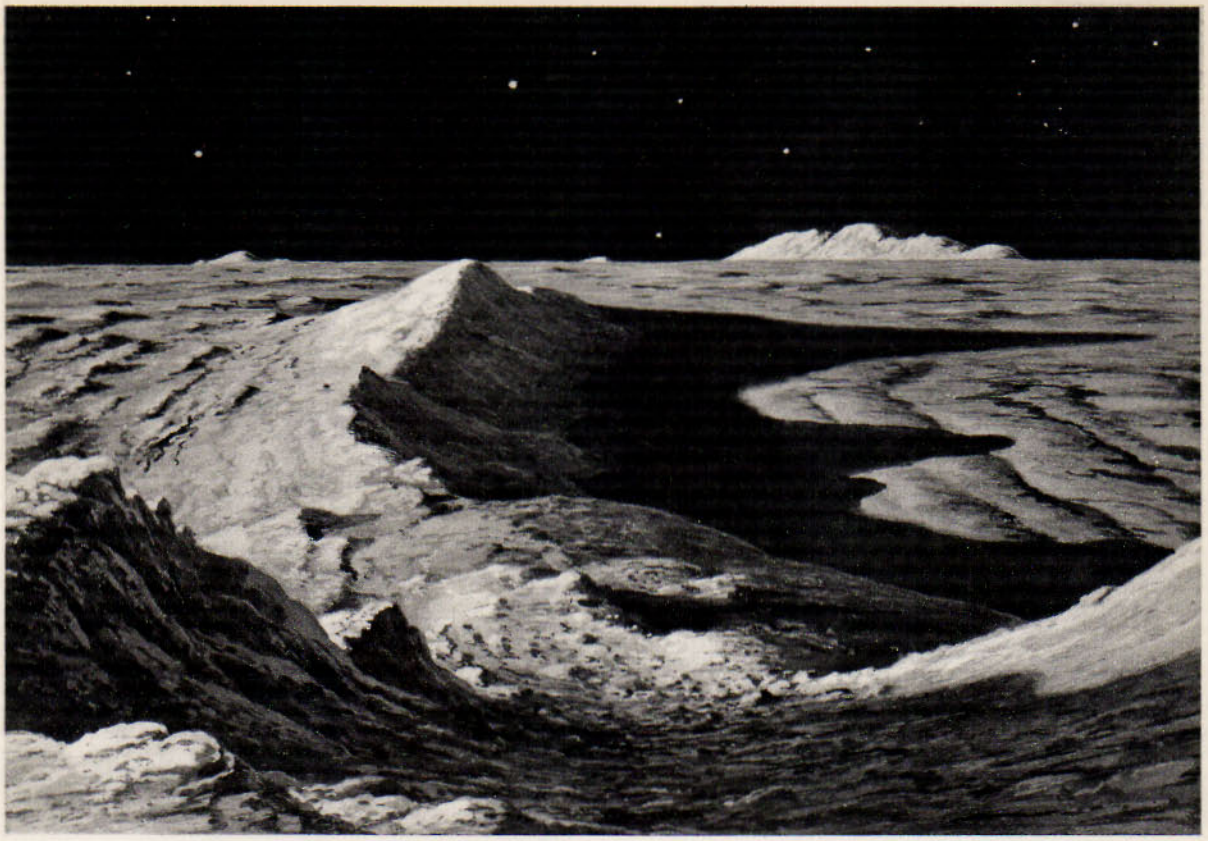


Fig. 299 — THE GENERAL NATURE OF THE LUNAR TERRAIN

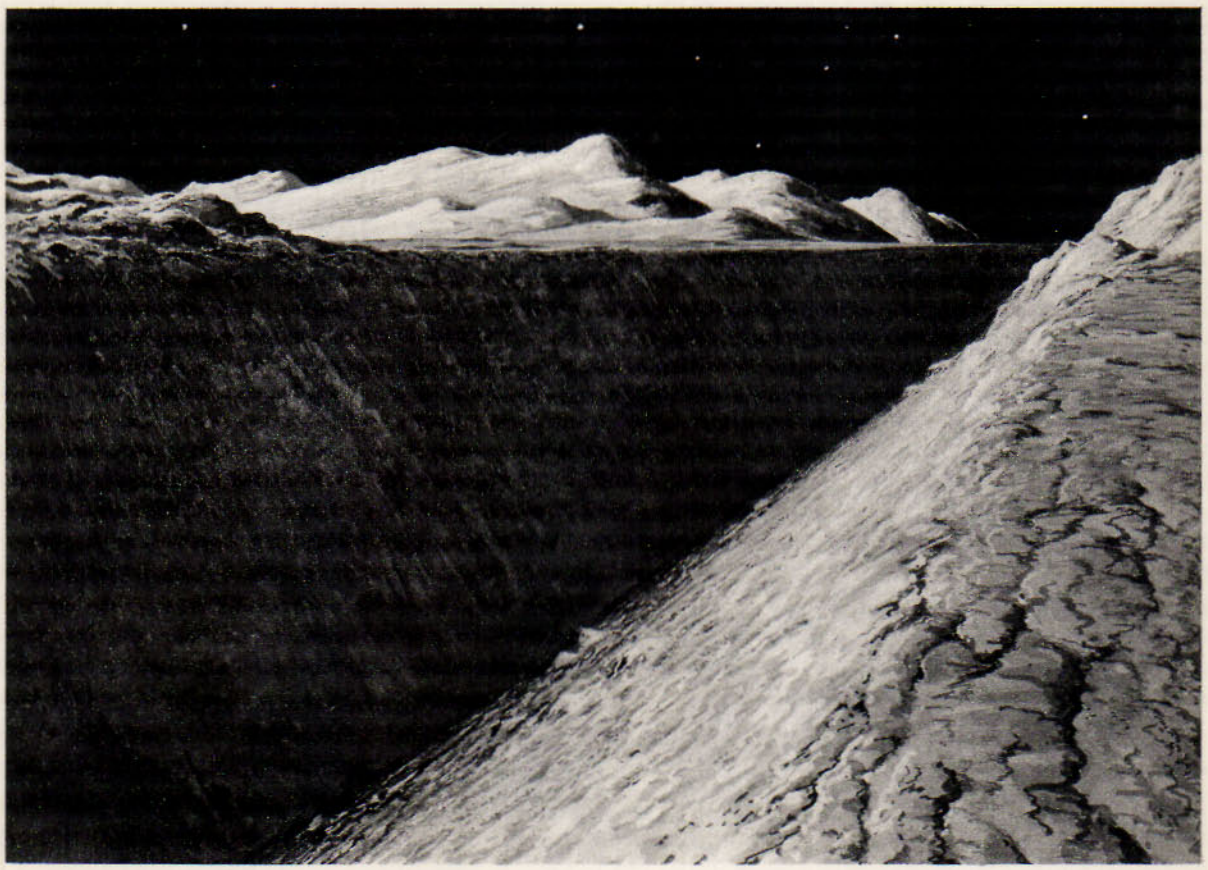


Fig. 300 — A LUNAR LANDSCAPE IN THE REGION OF A LARGE CREVASSE

For people on the Moon, our Earth would act as a wonderful astronomical time-keeper. By its phases the lunar day can be divided into various periods, and the rotation of the Earth can be utilised to give the subdivisions, because the familiar patterns of the oceans and continents regularly pass across the face of the disc. This, however is largely dependent on the state of the Earth's atmosphere, since there are at times so many cloud systems present that the actual surface must be hidden for days on end by sheets of thin haze, and it would be possible to pick out well-known landmarks only through gaps in these clouds. This constant change, however, would be very interesting to watch.

Although the Earth appears almost stationary from the Moon, the lunar sky seems to rotate in the same way as the Earth's, and all the heavenly bodies rise and set. Being the result of the Moon's orbital motion round the Earth, a complete rotation of the lunar sky takes 29 days, or about one month. Thus the stars first rise, then drift very slowly across the sky. If they are in the right position, they pass behind the disc of the Earth.

The Sun behaves in similar fashion, taking an hour to rise and set. There are no splendid displays of colour, as on Earth, since there is no atmosphere, but this is compensated for by the fact that a number of very interesting things can be observed which are difficult if not impossible to observe so well from our world. While the main part of the solar disc is hidden below the horizon, the pearly light of the corona stretches up into a pitch-black sky, to be followed by the brilliant chromosphere with its red, flame-like prominences. These are sights that we can only see on Earth for a few moments on the rare occasion of a total eclipse of the Sun. On the Moon they are visible to the fullest advantage (since there is no atmospheric absorption or scattering) every morning and evening.

Solar eclipses take place on the Moon when the Earth passes in front of the Sun. On Earth this is seen as a total eclipse of the Moon, since the latter is in the Earth's shadow. As seen from the surface of our satellite, however, the phenomenon has several interesting features not visible in an eclipse on Earth. For one thing the Sun disappears completely behind the large disc of the Earth; since the latter is four times as large as the Sun much of the spectacle is therefore hidden. One compensation is that the Earth is surrounded by a reddish ring of light: this is the Sun shining through the planet's atmosphere. We see this as a ruddy glow which suffuses the Moon's disc at the time of a total eclipse.

Such are the main visual impressions which would greet a human being being suddenly transported to the Moon, with apparatus allowing him to breathe, control his temperature, take food and drink and maintain life generally. What other physical sensations would he experience?

The harsh character of the landscape is aggravated by a terrible feeling of loneliness and desolation. No sound breaks the silence since there is no air to carry it. A whole mountain might topple, and it would have the dreamlike quality of a purely visual impression unless the shock waves reached the observer through the ground. The only sound, apart from the voices from his radio set, would be his own breathing.

The intense brilliance of the landscape would be made more trying by the fierce heat of the ground, which reaches a temperature of more than 100° C at noon; our traveller must be well protected by both filters and insulating materials, and must also have some sort of refrigerating apparatus. His troubles are increased by the excess of raw ultraviolet radiation coming from the Sun: long exposure to this would be very injurious if not fatal to a human being. Nor can he simply shelter from these hazards under cover of darkness or in the shadow of a mountain, for then he would be plunged into an unimaginable cold and would rapidly freeze to death unless his vacuum suit were provided with heaters. Once immune to all these dangers, however, he would find life made very interesting indeed by his extraordinary feeling of lightness due to the Moon's low gravity, which is only one-sixth as strong as the Earth's. A man weighing 180 pounds here would weigh only about 30 pounds on the Moon, and even a weakling would put to shame any Earth-bound weight-lifter by the load he could carry. Athletics would be child's play to a man who could jump over 15 or 20 foot obstacles without difficulty and travel for miles, without fatigue, in great leaps and bounds. Furthermore, the low gravity would reduce the dangers of falling: a man would have time to catch hold of something or twist to land on his feet since he would fall in what appeared to be slow motion. This would be of enormous advantage in reducing the difficulties and dangers in the exploration of the Moon, much of which would have to be carried out on foot owing to the broken nature of the ground, which is covered with mountains and crevasses.

This short account of the lunar world serves to show how different it is in many ways from its nearest neighbour in space, the Earth. One day men will surely conquer it, but it will be a hard struggle even to gain a momentary footing on this inhospitable world and no doubt there will be sacrifices to make. Apart from the lure of the unknown, the Moon will be a paradise for tourists — particularly for rock-climbers, who will be able to give lessons to the most agile chamois! Above all, for the richness of its unclouded and undisturbed skies, for the unmatched brilliance of the stars, the Moon will be a sort of paradise for astronomers, who will remove their instruments from the dusty, turbulent atmosphere of the Earth and set them up to observe the Universe from the 'Queen of Night', wondering how their forebears could ever have made as much progress in astronomy as they did from a place so unsuitable as the Earth!

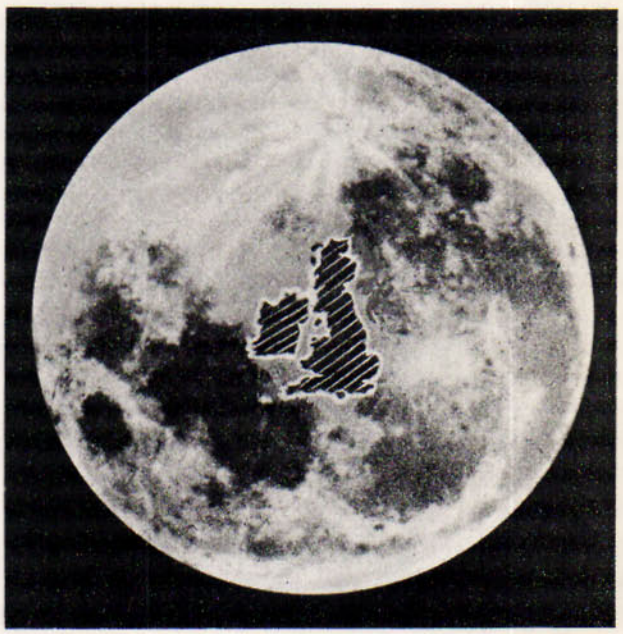


Fig. 301 — THE BRITISH ISLES ON THE SAME SCALE AS THE MOON

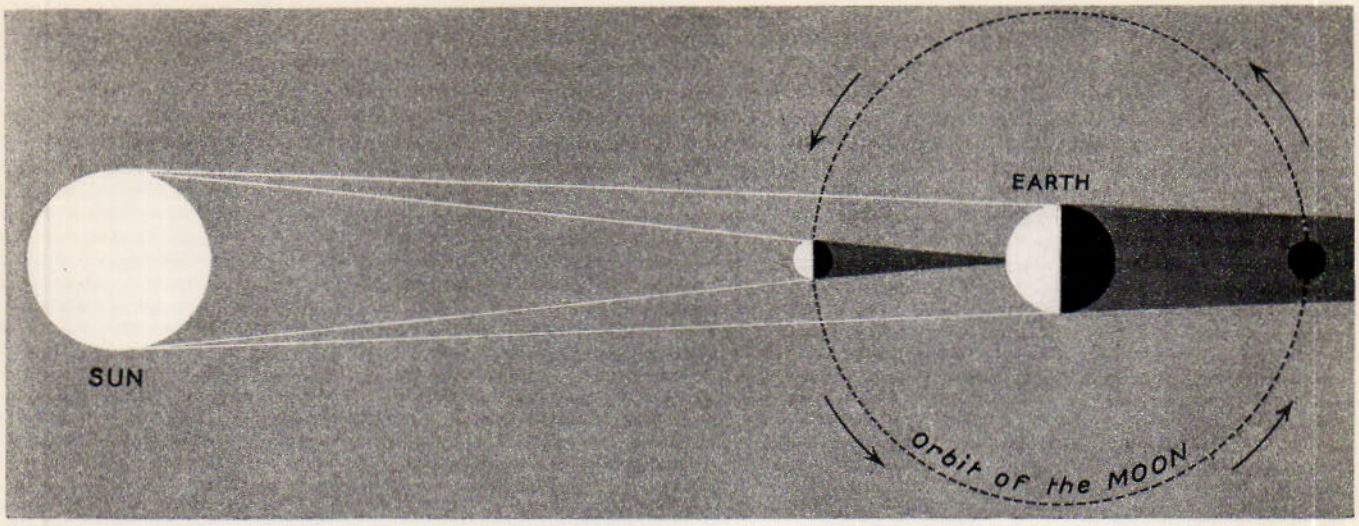


Fig. 302 — GENERAL MECHANISM OF SOLAR AND LUNAR ECLIPSES

CHAPTER 6

ECLIPSES, OCCULTATIONS AND TIDES

The Moon is not only the strange world we have just described; it is also a heavenly body in its own right, whose movement round the Earth and gravitational attraction cause phenomena of considerable interest to us human beings.

In an earlier chapter we discussed the complicated question of the Moon's motion, its ever-changing phases and its rôle as light-giver to the Earth when the Sun has set. Let us now examine two phenomena which result from the way in which our satellite moves: eclipses and tides.

An eclipse, as well as being a beautiful sight, is for the astronomer the source of vital information, without which it is doubtful whether astronomy could have advanced as

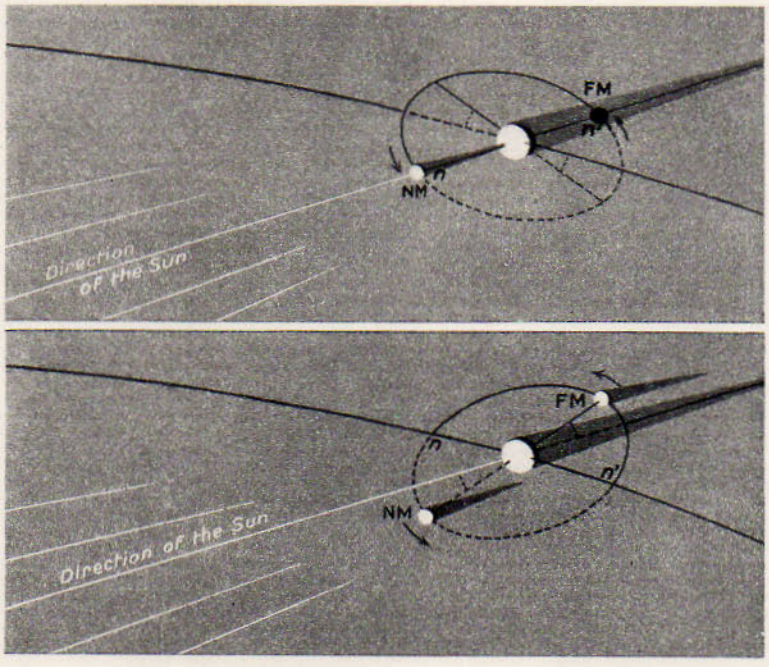


Fig. 303 — ECLIPSES CAN HAPPEN ONLY WHEN THE LINE OF THE NODES OF THE MOON'S ORBIT COINCIDES WITH THE DIRECTION OF THE SUN

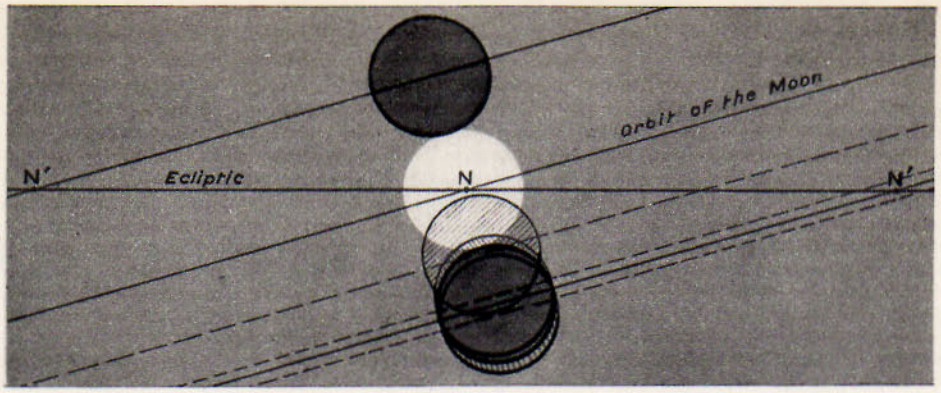


Fig. 304 — Because of the large apparent diameters of the Sun and Moon, solar eclipses can occur even when the line of the nodes points slightly away from the centre of the Sun's disc (between N and N').

far as it has. The tides are of practical importance for the inhabitants of the Earth, although they are not without a purely theoretical interest of their own.

ECLIPSES

The general mechanism of eclipses. — The word 'eclipse' is used to describe two quite distinct phenomena.

A solar eclipse occurs when the Moon passes directly in front of the Sun. An eclipse of the Moon occurs when our satellite passes behind the Earth into its shadow, and is thus no longer illuminated by the Sun's rays.

The general mechanism of eclipses has already been touched upon in the sections dealing with the Moon's motion, but the accompanying diagrams will bring the description back to mind. They show that the Moon's orbit actually turns in space with respect to the stars so that a different diameter of the orbit is always pointing sunwards. Eclipses can occur only when the line of the nodes coincides with this direction. This principle applies, strictly speaking, only to the centres of the Sun and Moon, and in practice the apparent diameters of these two bodies must also be taken into consideration.

To say that the line of the nodes points in the direction of the Sun does not necessarily mean that it passes through the centre of the solar disc, nor through the centre of the Earth on the other side. Because both the Sun and Moon have large diameters, it is enough, in a solar eclipse, for the line to point somewhere near the Sun, on either one side or the other of the line joining the centre of the Sun to the centre of the Earth; the Moon will then pass partly in front of the Sun (Fig. 304). In this case the eclipse is not total but partial. The same thing occurs in an eclipse of the Moon on account of the large diameter of the cone of shadow cast by the Earth.

The geometry of the situation is extremely simple and explains why solar eclipses are more numerous and frequent than lunar eclipses. Fig. 305 shows that for an eclipse to occur it is enough for the Moon to enter into the cone just containing the Sun and Earth: from A to B between Earth and Sun an eclipse of the Sun, and the passage A′ to B′ corresponds to an eclipse of the Moon in the shadow of the Earth. Now it is obvious that the Moon is more likely to pass from A to B than from A′ to B′, which is a narrower section; the difference is shown (exaggerated) by the length of the arrow AB, which gives the probability of eclipses of the Sun; on the other side, much of the arrow lies outside the cone of shadow and the probability of a lunar eclipse is therefore less. From this we find that eclipses of the Sun are more frequent than eclipses of the Moon in the ratio 4:3.

The total number of eclipses per year is not constant. The maximum is 7 (4 or 5 solar and 3 or 2 lunar), the minimum is 2 (which are always solar). What is the reason for this irregularity in number and frequency?

It arises from the fact that eclipses are the result of

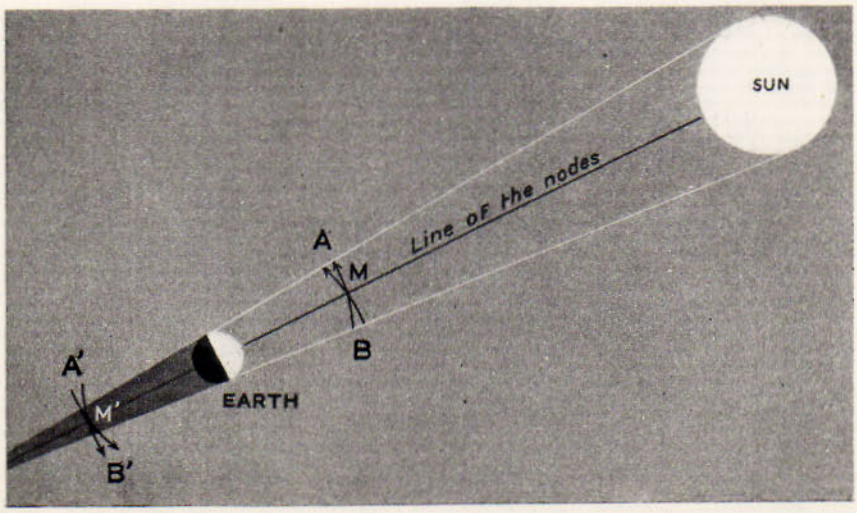


Fig. 305 — FREQUENCY OF SOLAR AND LUNAR ECLIPSES

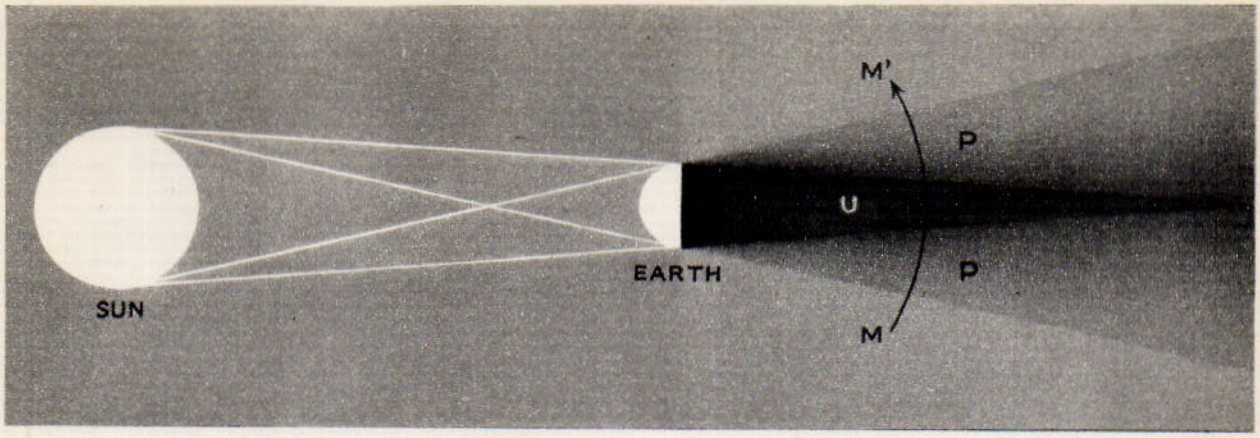


Fig. 306 – ECLIPSES OF THE MOON. While moving from M to M', the Moon passes successively into the cones of penumbra and umbra (P and U respectively).

a very complex set of movements: the orbital motion of the Moon, which is subject to variations, and the revolution of the Earth which moves the whole system round the Sun. As a result of these and other causes, the required conditions occur at irregular intervals, which are principally determined by the most important of the contributing motions: the rotation of the lunar orbit, which causes the nodes to regress. Broadly the same sequence of conditions is repeated at every complete revolution of the Moon's orbit, so that eclipses in similar circumstances are repeated after the same interval.

Allowing for minor variations, the complete cycle takes place in about 18 years 11 days; this cycle, which was known even to the early astronomers of Mesopotamia, is called the Saros. During this period, which is equal to 223 lunations, there are on average 71 eclipses, 43 of the Sun and 28 of the Moon.

The Saros, however, is an approximate period serving to date the return of an eclipse but not as a perfectly accurate means of prediction. Let us consider the case of a partial, almost total eclipse occurring on a certain date. After 18 years 11 days, this eclipse will occur again but this time with a slight difference one way or another. It may be either total or partial, but if it is still partial, the amount of the Sun's disc covered will not be the same as before. And it will not be visible in precisely the same part of the world.

The reasons for these differences will be made clear in the detailed examination we are going to make of the two categories of eclipses, and the conditions under which they occur. Because an eclipse of the Moon is due to the cutting off of the Sun's light, it is observable from the entire night hemisphere of the Earth. Thus for any one country, the visibility of the eclipse depends on the hour at which in occurs: so long as this is between the time of moonrise and moonset, the eclipse will be observable. The Sun, on the other hand, is eclipsed because the Moon passes in front of it and casts a shadow on the Earth. Owing to the large relative size of the Earth, effects of perspective may well, according to the observer's position on the globe, mean that the Moon passes either above or below the Sun, and so an eclipse which is total at one place is only partial at another, and diminishes progressively until no part of the Sun is obscured at all. For example, in 1956 there was a total eclipse in Sweden which was only partial in Britain. The time of day is also important in this connection. The problem is further complicated by the complexity of the Moon's movements and their way of combining with the Earth's rotation about an inclined axis. This means that some eclipses are visible only around one of the poles, whilst others are clearly seen from the tropics.

Eclipses of the Moon. – Now that we know the fundamental conditions causing eclipses, let us examine the phenomena in more detail.

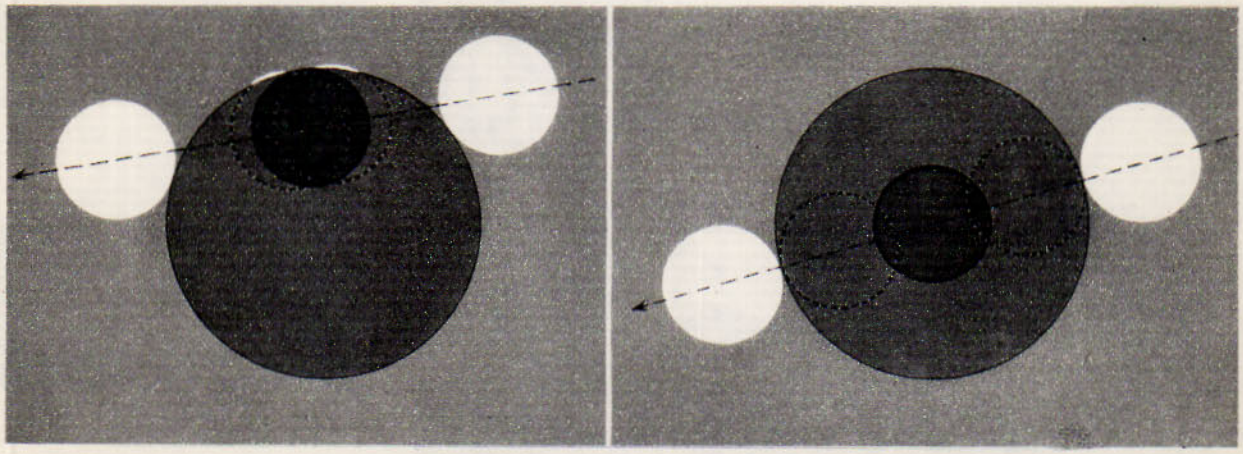


Fig. 307 – A TOTAL ECLIPSE OF THE MOON. Left: the Moon is just immersed in the Earth's shadow; right: the Moon passes exactly through the centre of the shadow.

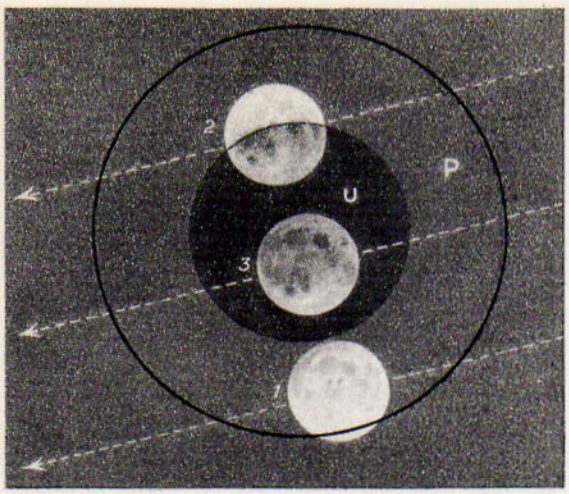


Fig. 308 — DIFFERENT SORTS OF LUNAR ECLIPSE IN THE CONES OF PENUMBRA (P) AND UMBRA (U). 1, eclipse by the penumbra; 2, partial eclipse; 3, total eclipse.

As shown in Fig. 302, the Earth, which is illuminated by the Sun, casts a cone of shadow away from the latter. The shape of this cone is defined by tangents from the outside of the terrestrial globe to the edge of the Sun, which is much the larger. From the respective diameters of these two bodies and their mean distance apart, it can be calculated that the length of the shadow-cone is 108 times the Earth's diameter. As the Moon's distance is only 30 Earth diameters, it follows that our satellite will pass through the cone of shadow where this cone is still fairly wide (as shown by the arrow in Fig. 306).

Moreover the umbra, the darkest part of the shadow, is surrounded by a cone of penumbra, which broadens outwards from the Earth. Its shape is determined by the interior tangents between the solar and terrestrial globes. This means that the Moon must pass through the penumbra both before and after passing through the umbra itself.

The angle at which the Moon passes through the Earth's shadow is determined by the orientation of its orbit at the time, and this also controls the magnitude of the eclipses — whether partial or total — and the succession of visual effects which may be observed.

As mentioned earlier, if the line of the nodes is pointing exactly in the Earth-Sun direction, the Moon's path will pass right through the centre of the shadow. If the line of the nodes is not perfectly orientated in this direction, then the Moon will pass either a little above or below the centre. But so long as our satellite's path cuts the penumbra, some sort of eclipse will take place.

When passing through the penumbra, the Moon is receiving less light than usual, for the Sun is then partially obscured by the Earth. This illumination continues to decrease the nearer the Moon comes to the umbra proper, and only when the Moon is completely immersed is the eclipse said to be total. The accompanying diagrams clearly show these various stages.

According to the orientation of the nodes, therefore, eclipses may be of three types:

(1) Penumbral eclipse, when the Moon passes only through partial shadow and comes out again without touching the umbra;

(2) Partial eclipse, when the Moon passes through the

Fig. 309 — ECLIPSE OF THE MOON: entry into the penumbra and umbra.

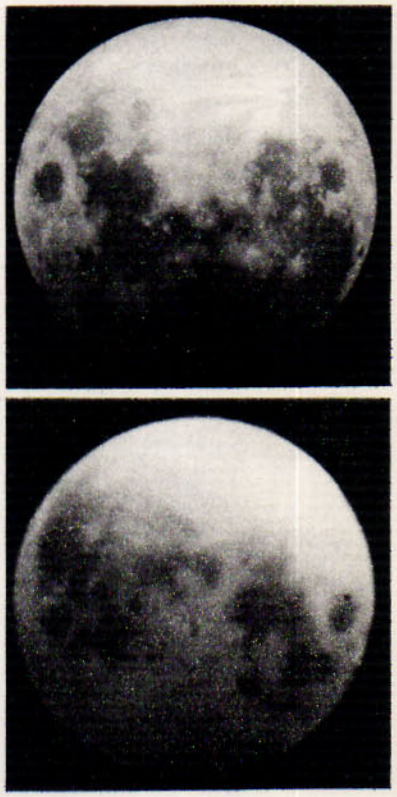


Fig. 310 — ABOVE: THE MOON IN THE PENUMBRA; BELOW: THE MOON DURING A TOTAL ECLIPSE. Photographed in infra-red light.

Fig. 311 — PROGRESS OF A TOTAL ECLIPSE, from the entry into the umbra to the moment of emergence.

penumbra but only partly through the umbra before emerging;

(3) Total eclipse, when the whole Moon passes through the umbra.

Fig. 308 shows these three cases in diagrammatic form, 2 and 3 being eclipses proper. The various phases of eclipses are given in almanacs in the following way:

(a) Entry into the penumbra, greatest phase at mid-eclipse, and emergence from penumbra;

(b) Entry into the penumbra, entry into the umbra, mid-eclipse, emergence from umbra, emergence from penumbra;

(c) Entry into penumbra, entry into umbra, commencement of totality, mid-eclipse, end of totality, emergence from umbra, emergence from penumbra.

These elements are variable, since the angle at which the Moon passes through the Earth's shadow changes according to the orientation of its orbit, thus producing the different types of eclipses. Moreover, the varying distance between the Sun and Earth affects the length of the cone of umbra and penumbra, and, even more important, the varying distance of the Moon, which alters its apparent diameter, means that it will pass through the shadow cone further from or nearer to the Earth, therefore at a wider or narrower section of the cone than usual.

The maximum duration of a lunar eclipse is 2 hours for passing through the umbra, and 4 hours through both the umbra and the penumbra.

The brightness of the eclipsed Moon is both variable and unpredictable. Sometimes it disappears altogether, but it can usually be seen even at mid-eclipse shining with a coppery glow. There is of course a gradual diminution in brightness as the Moon passes through the various states to totality, and the illumination at some of these states is shown in Fig. 311.

The coppery glow mentioned above is an interesting phenomenon. By rights the Moon ought to disappear entirely when totally eclipsed, since it is cut off from the rays of the Sun. However, at most eclipses the Moon is still bright enough for most of the main surface features to be visible.

The reason for this lies in the presence of the Earth's atmosphere. We have already seen that the Earth's shadow is a cone, its base corresponding to the terrestrial diameter. Rays of light just grazing the edge of the globe must pass right through the thickest part of the atmosphere, whereupon they undergo refraction, are bent into the cone of shadow and fall upon the eclipsed Moon (Fig. 312). Moreover these rays also undergo absorption in the atmosphere, the lower their path the greater being the amount of absorption, and consequently the greater the loss to the blue and violet end of the spectrum. Exactly the same thing, it will be remembered, occurs at sunset. Thus the eclipsed Moon takes on a coppery or blood-red colour, according to conditions in the Earth's atmosphere. As will be seen from the illustration, the edges of the disc are lighter than the centre owing to the path taken by the light-rays.

In fact there is a considerable range in the brightness and colour of the eclipsed Moon; sometimes it is a bright orange-yellow, sometimes a dull red, sometimes in between, according to how centrally it cuts through the shadow-cone.

Words cannot convey the beauty of a total eclipse of the Moon, nor the strange impression produced by the

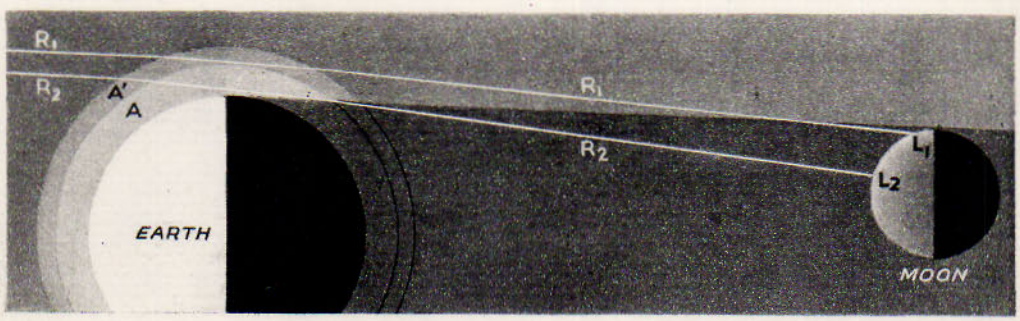


Fig. 312 — THE ILLUMINATION OF THE MOON'S DISC DURING AN ECLIPSE. The ray of light R_2, which passes through the greatest thickness of the atmosphere, is more strongly refracted than R_1, which passes only through its upper regions.

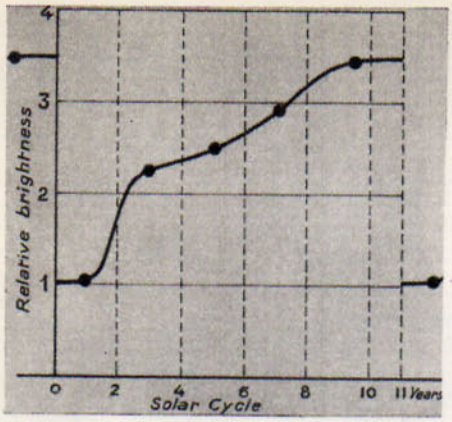


Fig. 313 — BRIGHTNESS OF TOTAL ECLIPSES OF THE MOON THROUGH THE COURSE OF A CYCLE OF SOLAR ACTIVITY *(after A. Danjon)*

sight of a dull red disc apparently bathed in the glow of some great celestial inferno, set in a sky filled with stars which the normal brilliance of the full moon would prevent our seeing.

As we have seen, no two eclipses are alike, since there is always some factor which may cause variation in the phase, brightness or colour; consequently it is worth while observing them merely on this account. However, much useful information may be gained from telescopic observation of the eclipsed Moon.

Information derived from lunar eclipses. — The often considerable differences between the circumstances of two eclipses are not really to be wondered at, because in order to reach the eclipsed Moon, rays of light must pass through the essentially changeable medium of the Earth's atmosphere, and a study of the various effects produced on our satellite may give us valuable information about conditions existing at various heights in our own atmosphere. This information may corroborate or complement data derived from more localised sounding methods.

We have already seen how rays of light are refracted through the atmosphere on to the eclipsed Moon. From the known value of refraction, it is possible to calculate at what height the rays of light illuminating a given part of the Moon's surface have passed through the terrestrial atmosphere. Spectroscopic analysis (see Chapter 20) enables us to determine the composition of a gaseous medium through which light rays have passed; an examination of various parts of the eclipsed Moon will therefore tell us the properties of the atmospheric layers in question.

As an example, consider the work done by Barbier, Chalonge and Vigroux at Paris Observatory on the occasion of the total eclipse on 2 March 1942. Photographs of the Moon's spectrum clearly revealed the unequal distribution (in height above the Earth's surface) of oxygen and ozone. They thus confirmed that ozone is particularly plentiful in the stratosphere at heights of between 10 and 20 miles approximately. Further work on these lines, with increasingly refined instruments and techniques, will no doubt provide a great deal more accurate and complete information about the composition of the higher regions of the atmosphere.

Let us now study the facts to which A. Danjon has drawn attention and which are of quite special significance. Here it will to some extent be necessary to anticipate later chapters on the Sun, whose radiation of heat and light directly or indirectly controls conditions obtaining on the various planets. The Sun is the seat of violent activity, the intensity of which fluctuates over a period of 11.1 years (which is the figure we met in our discussion of terrestrial magnetism).

Danjon studied a long series of total eclipses of the Moon, and his work shows that the intensity of the illumination and the reddish colour of the disc may be correlated with the solar cycle.

The relation is as follows:

(i) In the two years following a minimum of solar activity, the Moon is only feebly illuminated and only slightly coloured when in the Earth's shadow.

(ii) As the minimum is left behind, eclipses become much brighter and redder. In the three or four years immediately preceding the next minimum it is very bright, and of an orange-yellow colour.

The maximum phase of the solar cycle passes completely unmarked, whilst the minimum phase is indicated by a sudden and considerable diminution in brightness and colour; this change forms a discontinuity.

The accompanying graph shows how the luminosity curve rises between two minima, with a sudden fall as it reaches the actual minimum. It should be noted that the coefficients of brightness used as ordinates are on an arbitrary scale, derived, since precise photometric

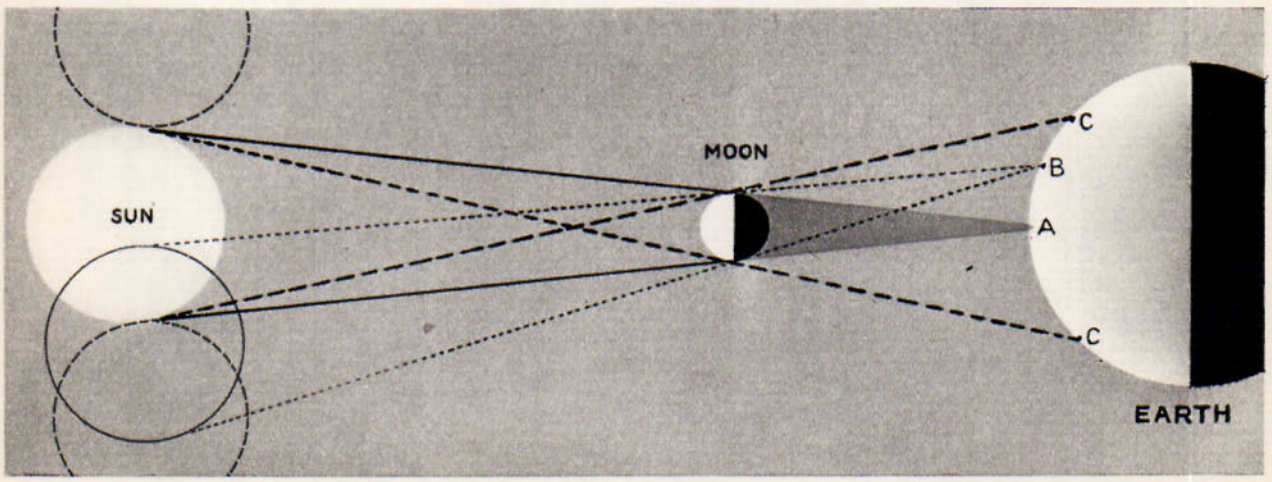


Fig. 314 — PRINCIPLE OF PARTIAL AND TOTAL ECLIPSES OF THE SUN

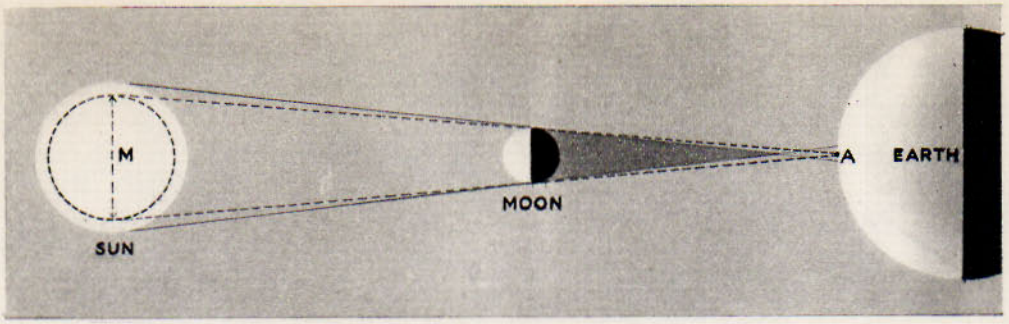


Fig. 315 – PRINCIPLE OF ANNULAR ECLIPSES OF THE SUN

observations are not available, from eye-witness acounts of old eclipses. The scale of brightness is given below for the benefit of future observers.

0 Very dark eclipse. Moon practically invisible at mid-eclipse.
1 Dark grey or brownish eclipse. Surface features difficult to make out.
2 Dark red or ruddy eclipse, frequently with a large dark patch at the centre of the umbra, surrounded by a slightly brighter outer zone.
3 Brick-red eclipse. Umbra frequently bordered with a grey or bright yellowish zone.
4 Coppery red or orange eclipse, very bright. Luminous outer zone of a bluish tint.

There have been exceptions to the general run of the series which merit study. For example in the series of eclipses at the end of the last century, those of 4 October 1884, 30 March and 24 September 1888 were very dark, whereas they ought to have been bright according to their position in the solar cycle. This abnormality is quite consistent with the abnormal state of the atmosphere at the time, for it must be remembered that the eruption of Krakatoa took place in August 1883, and the dust and volcanic ash then expelled into the atmosphere were detectable for two or three years after the catastrophe.

Similar effects were observed following various eruptions in 1902–3, and in 1913, after that of Katmai, Alaska.

Thus it will be realised that eclipses of the Moon are occurrences of great interest and they should be observed as frequently as possible; they not only give us valuable information about the Earth's atmosphere but also pose the wider problem of how exactly the solar cycle influences the behaviour of that atmosphere.

Eclipses of the Sun. – The theory behind eclipses of the Sun is as simple as that of lunar eclipses, but the actual circumstances are far more variable and this adds extra complications – for example, an eclipse of the Sun is visible from only a very small part of the Earth's surface.

The most important factor in the production of a solar eclipse is the relative dimensions and separation of the three bodies involved, the Sun, Moon and Earth.

As shown in Fig. 314, the Sun is enormous but very far away, and the Moon is small and near; the tangents between these two globes define the shape of the thin cone of shadow cast by the Moon. The narrowest point in this shadow is situated somewhere near the Earth's surface. Thus at A, where the point of the shadow-cone falls on the globe, it can be seen that if the Moon's apparent diameter is exactly the same as the Sun's, the Moon will totally eclipse the Sun, whilst tangents to the edge of the lunar disc from elsewhere on the Earth (B) show that the Sun is only partially eclipsed, and from C, which is even further away, the Moon does not cover any part of the Sun and there is consequently no eclipse at all.

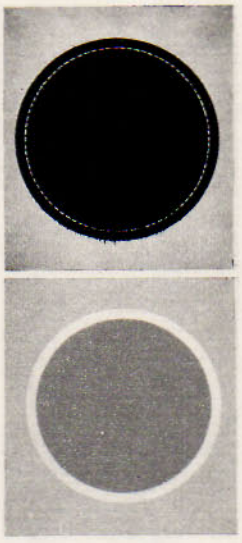


Fig. 316 – GREATEST MAGNITUDE OF TOTAL ECLIPSES (ABOVE) AND ANNULAR ECLIPSES (BELOW)

This is only intended to demonstrate the principle of eclipses, but it will help us to understand the details better. In the first place, the exaggerated proportions of the three bodies in the diagram show that at any particular moment the Moon is not at the same distance from every

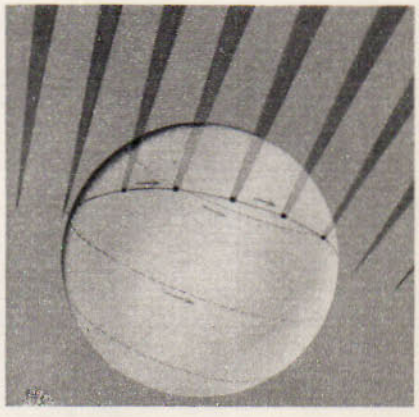


Fig. 317 – MOVEMENT OF THE MOON'S SHADOW OVER THE TERRESTRIAL GLOBE

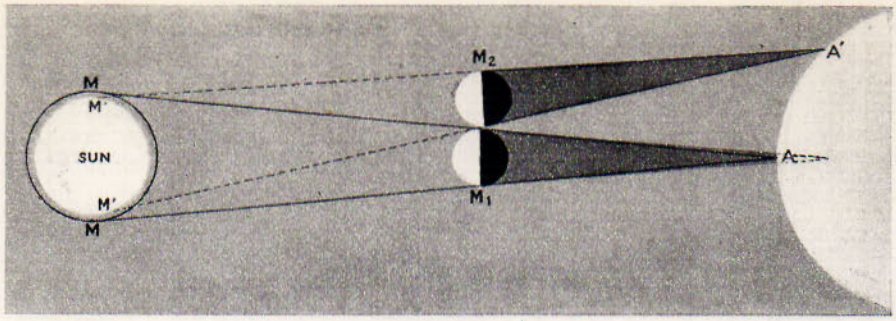


Fig. 318 — DIFFERENCE IN THE SIZE OF THE SAME TOTAL ECLIPSE OF THE SUN seen from two widely separated points on the surface of the Earth. The apparent diameter of the Moon is greater when seen from A at M_1 than from A′ at M_2.

point on the Earth's surface, which is to say that its apparent diameter varies slightly over the surface of the globe, thus altering the magnitude of the eclipse.

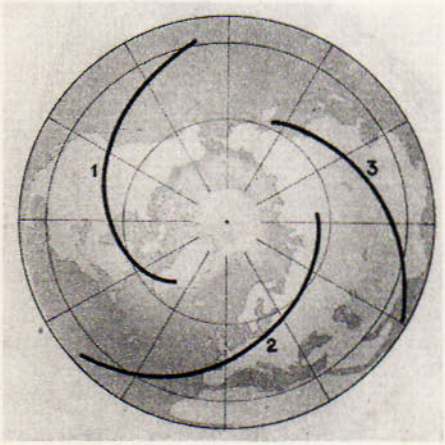


Fig. 319 — THE SAME ECLIPSE OF THE SUN, WHEN REPEATED AFTER THE 18-YEAR 'SAROS' CYCLE, DOES NOT OCCUR OVER THE SAME PARTS OF THE EARTH, BUT SUCCESSIVELY AT 1, 2 AND 3

Herein lies the great importance of the inherent complexities in the Moon's distance from Earth and its orbital motion, and the Earth's rotation about its inclined axis. We shall consider these factors in order.

We have just seen how the cone of shadow is established and causes a total eclipse of the Sun according to the distances and dimensions of the three bodies. If, however, the Moon is at apogee (Fig. 315) the point of the cone no longer reaches the Earth's surface. For an observer at A, the Moon's apparent diameter will be smaller than the Sun's (the dotted line) and there will be a ring of the Sun showing behind the Moon at mid-eclipse: for this reason it is called an annular (ring-shaped) eclipse. The size of the ring will vary according to how near or far the Moon is from the Earth at the time; when the Moon is furthest away, the area of the Sun not eclipsed will be greatest. When the Moon is so placed that the point of the shadow-cone just touches the Earth, there will once more be a total eclipse.

Conversely, when the Moon is at perigee, the length of the shadow-cone is greater than the distance between the Earth and the Moon so that there will be a circular zone on the surface where a total eclipse is visible. Since the Moon then looks appreciably larger than the Sun, for several minutes the Moon will cover the Sun completely, the maximum duration of eclipse being enjoyed by observers stationed at the centre of the zone of totality. Between these two extremes, every intermediate stage is possible, but the phenomenon is made more complicated by the following facts.

In the first place, because of the Moon's orbital motion, its cone of shadow sweeps across the Earth's surface (Fig. 317) and the line thus traced out is called the belt of totality. But the ever-changing orientation of the lunar orbit, coupled with the Earth's rotation and position in relation to the Sun, means that these belts of totality do not always cross the same parts of the Earth's surface:

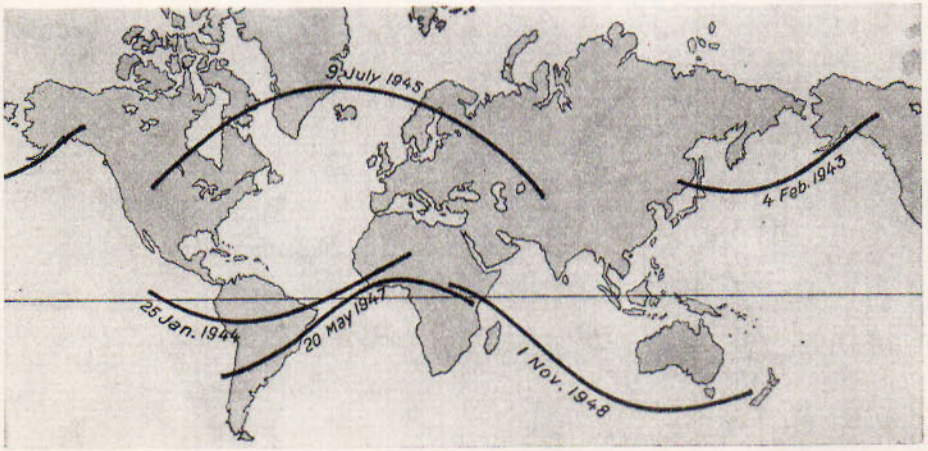


Fig. 320 — DISTRIBUTION OF THE BELTS OF TOTALITY FROM 1943 TO 1948

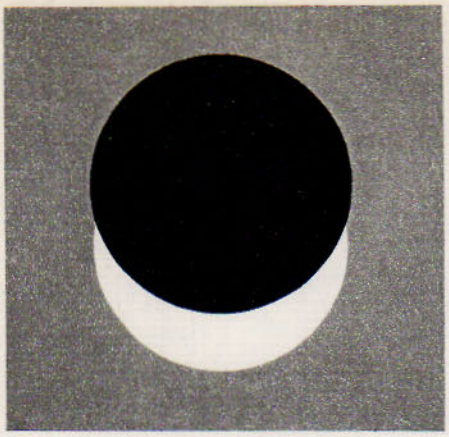


Fig. 321 — MAXIMUM PHASE IN LONDON OF THE ECLIPSE OF 30 JUNE 1954, WHICH WAS TOTAL IN SWEDEN

sometimes they pass across high latitudes, sometimes they keep to the tropics. (Generally speaking, however, they are in the habit of passing over the most inaccessible and uninhabited regions of the Earth much more readily than over places where the eclipse can be viewed with ease by the greatest number of people!) The spherical shape of the Earth must also be considered because different places are at different distances from the Moon. Figs. 317 and 318 show the differences resulting from these inequalities. Thus a single eclipse may be of different magnitudes at various points along its track: at one place it may be annular, whilst farther on, where the Earth's surface is slightly nearer the Moon, it may be total.

On either side of the belt of totality a partial eclipse will be seen. The width of this zone is determined by the relative positions of the three bodies, and the magnitude of the eclipse decreases outwards from the belt of totality. Lastly, it may happen that because the node in question is too far from the centre of the Sun, the cone of shadow will pass a little above or below the Earth, thus producing a partial eclipse in the polar regions.

Up to now we have considered eclipses in relation to the terrestrial globe without allowing for the fact that the Earth is rotating. As shown earlier, the Earth's rotation means that the movement of different parts of the planet's surface in relation to the Moon's shadow will not be uniform, and Fig. 317 shows the complicated path which results from the way in which the two movements combine. Although these are both in the same direction, they are not equal in speed and are not parallel, owing to the respective orientations of the Moon's orbit, the Earth's surface and the axis of rotation relative to the Sun. The latter of course depends on the season.

All these elements must be allowed for in order to predict a solar eclipse accurately: in which parts of the Earth it will be seen, whether it will be partial, annular or total and exactly at what time it will occur, whether just as the Sun is rising or setting at a particular place. The example given in Fig. 320 shows just how complex this task can be.

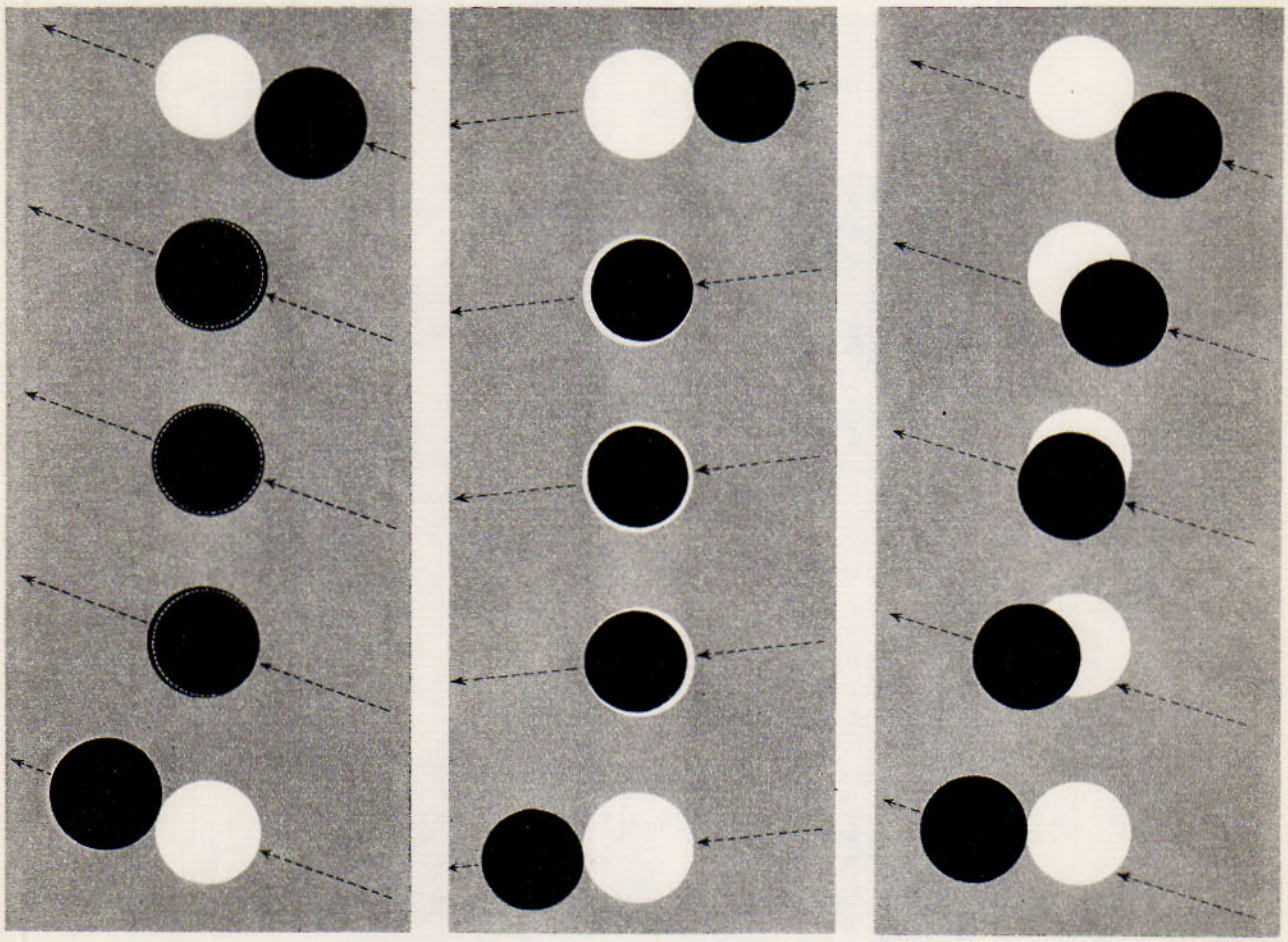


Fig. 322 — SUCCESSIVE PHASES OF AN ECLIPSE OF THE SUN

Total eclipse Annular eclipse Partial eclipse

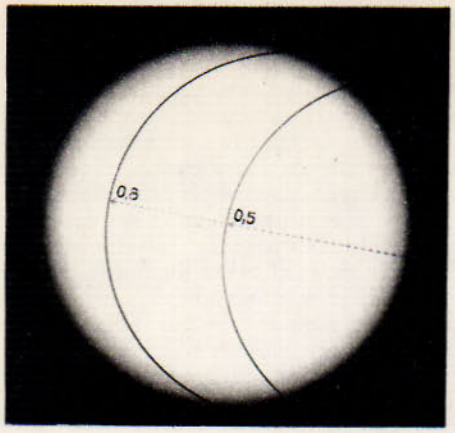


Fig. 323 — MAGNITUDE OF A SOLAR ECLIPSE

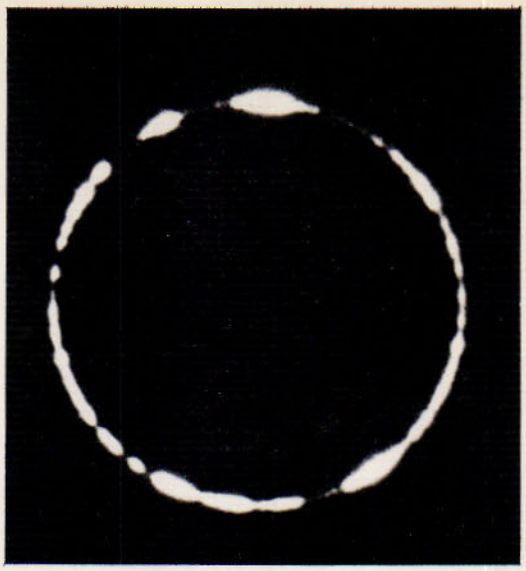


Fig. 325 — TOTAL ECLIPSE, SHOWING 'BAILY'S BEADS'

Visibility of solar eclipses. — As we saw in the preceding section, eclipses of the Sun, unlike those of the Moon, are visible only from a limited part of the Earth's surface, and different eclipses are seen from very widely scattered parts of the world. Even two eclipses in the same position in the Saros cycle do not occur at the same spot on the globe because of the Earth's rotation (Fig. 319).

It follows that for any particular place, the occurrence of an eclipse, and especially of any particular sort of eclipse — total, partial or annular — is extremely rare. In the neighbourhood of Paris, for example, the last total eclipse took place on 17 April 1912; the next one will be on 11 August 1999. The people of Paris will have to wait longer still for the next one: it will be on 23 September 2090, and the succeeding one will be on 4 June 2160 . . .

On the other hand, partial eclipses are much more frequent, since it is sufficient for the spot to be somewhere

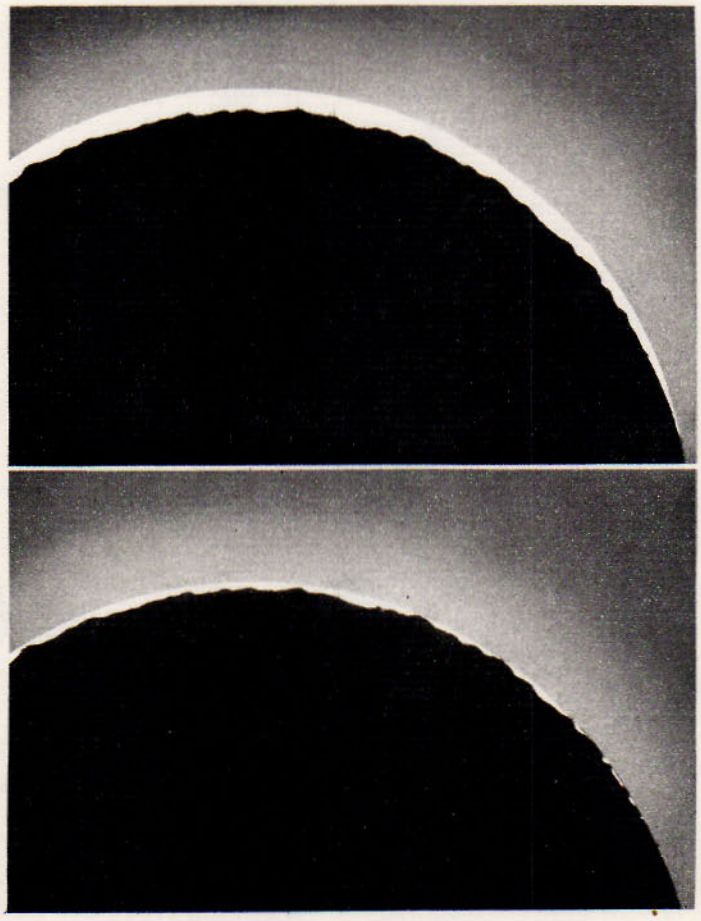


Fig. 324 — FORMATION OF 'BAILY'S BEADS'

Fig. 326 — (a) A MOUNTAIN LANDSCAPE AS NORMALLY ILLUMINATED BY THE SUN

(b) THE SAME, PHOTOGRAPHED DURING A TOTAL SOLAR ECLIPSE

in the penumbral belt, which is much wider than the belt of totality and therefore gives an increased chance of seeing an eclipse. It follows from this that the magnitude of an eclipse may vary considerably when seen from points which are not very far apart. This is due to perspective. The magnitude of the eclipse will depend on whether the observer is near the centre of the belt of totality or is right on the edge of the shadow. For example, an eclipse that was total in Sweden on 30 June 1954, was only partial in Britain, which was a few hundred miles from the centre of the totality belt (Fig. 321).

It is always very interesting to watch the progress of an eclipse, as, amongst other things, it provides a good demonstration of the Moon's motion as it creeps in front of the Sun. Moreover it forms an excellent practical lesson in astronomy, for which telescopic aid is not required. Indeed, so much happens all around in so short a time that a telescope may be more of nuisance than it is worth. There is one feature of an eclipse which is specially worth noticing: observe the little patches of light in the shadow of a tree during the eclipse. These patches of light are in fact pin-hole images of the Sun and they will show all the phases of the eclipse in miniature. The same effect can be seen more clearly when a little pencil of light enters a darkened room through a gap in the blind.

Fig. 322 shows the various phases of an eclipse and also the progressive variations in magnitude and direction. Comparison is made between the principal types of eclipse – total, annular and partial – and it is possible to see the principal phases which are given in ephemerides, i.e. the contacts between the edges of the Sun and Moon. The size of an eclipse is defined as the magnitude of the maximum phase. For partial eclipses, this magnitude is expressed as the fraction of the Sun's diameter hidden by the Moon (Fig. 323). Thus an eclipse of magnitude 0.5 means that at mid-eclipse, half the Sun's diameter is covered by the Moon. In the case of an annular eclipse, the magnitude is still expressed as a fraction, but the Moon is considered to be concentric with the Sun's disc, which is visible all round it.

The Moon's velocity, size and position relative to the Earth and Sun mean that it can cover the Sun completely for a maximum duration of 7m 58s at the Earth's equator and 6m 10s for a latitude of approximately 50°. In an annular eclipse, the Moon can remain completely in front of the Sun for a maximum of 12m 24s at the equator and 9m 56s at latitude 50°.

There is however, a special case, intermediate between the last two; this is the 'pearl ring' eclipse. Up to now we have only considered eclipses geometrically, taking the Moon, for example, as a perfectly smooth sphere. But we already know how rough the surface of the Moon really is, and consequently how uneven this makes the edge of the disc. Therefore, if the phase of an eclipse is such that it is *just* total, the edge of the Sun will still be visible because of occasional depressions in the surface of the Moon, which let the light through. Much the same thing may occur in an annular eclipse when the Sun is almost completely but not quite covered by the Moon. Mountains projecting from the edge of the Moon may break up the very thin ring of the Sun which should theoretically remain visible. This gives the appearance of a number of brilliant points of light round the edge of the Moon, known as '*Baily's Beads*'. This appearance is almost always seen at a total eclipse just as the Moon's limb is about to cut off the last thin segment of Sun. In the above cases, the eclipse may be described as either annular or total, or as both at the same time. This was true for the 1912 eclipse visible from Paris.

Before leaving the subject of eclipses, one possible misunderstanding must be cleared up. In our illustrations the Moon is shown quite clearly as a black disc cutting off all or part of the Sun. However, what we in fact see is that an ever-growing segment of the Sun is being blackened out for the Moon's disc is rendered

Fig. 327 – PARTIAL ECLIPSE OF THE SUN

Plate II. TOTAL ECLIPSE OF THE MOON

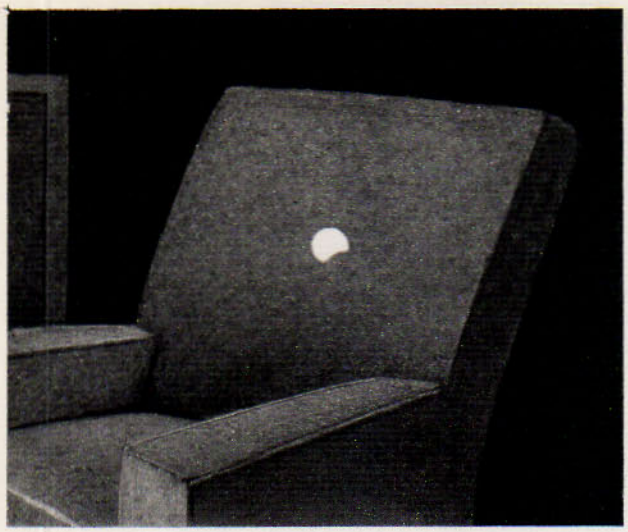


Fig. 328 — IMAGE OF A SOLAR ECLIPSE PROJECTED ON TO A CHAIR THROUGH A SMALL CRACK IN A SHUTTER

invisible, as at new moon, by the dazzling brilliance of the solar disc. As the eclipse progresses, the broad crescent, which is all that can be seen of the Sun, becomes smaller and smaller until it disappears at the moment of totality. Then the sky, having become dimmer and dimmer, suddenly goes dark and no longer hinders the eye. At this moment, the corona, which is usually lost in the general brightness of the sky, begins to stand out in all its splendour as a pearly sheen in the darkness, surrounding the Moon's disc which looks black by contrast. No words, drawing or photograph can really convey the breath-taking beauty and 'atmosphere' of this moment which seems to affect the whole of nature, and it is indeed a pity that so few people ever have the opportunity of seeing a total eclipse of the Sun. It is something never to be missed.

Generally speaking, the illumination decreases from the beginning of the eclipse onward, but this diminution is not usually very noticeable until about half of the Sun has been covered. When the magnitude of the eclipse has reached 0.8, the sunlight takes on a different tint since the thin crescent from which it is coming is darker and redder than the rest of the Sun's disc. From now on, the light fails with seemingly catastrophic rapidity.

Leaving aside the superstitious terror which grips many primitive people when the Sun so mysteriously disappears and plunges them into night in the middle of the day, we should note the characteristic effect an eclipse has on the animal and vegetable kingdoms. Farmyard animals crouch together or retire to the shelters where they spend the night, while all the birds stop singing and behave as they do at sunset. Bats, owls, insects and other nocturnal creatures emerge from their holes to hunt and be hunted in the darkness. Insects, indeed, seem to be deeply affected by the sudden darkness: ants abandon their labours but recommence as soon as the Sun reappears. Higher animals often show signs of fear; horses, mules and oxen refuse to move, and cower to the ground. Hungry dogs have been known to stop feeding at the onset of totality, starting again only when the Sun has returned. Certain plants and flowers which close up at night will also close up during a total eclipse, especially if it is of long duration.

These are the more spectacular phenomena from the layman's point of view, but astronomers have other kinds of information to gather.

The use of solar eclipses. — When the Moon passes in front of the Sun, it cuts off the solar radiation — heat, light, etc. — from the area over which the shadow passes, and this inevitably means that there will be some disturbance of the normal equilibrium. The most pronounced of these variations is the change in temperature, which in its turn sets off others. The electrical properties and the conductivity of the atmosphere are also affected to a certain extent; furthermore, there are repercussions in the state of ionisation of the upper atmosphere, which means that the propagation of radio-waves will be affected. At the moment we only have space to touch upon the numerous and delicate problems which interest the specialist and we shall naturally concentrate on the astronomical side of the question, since eclipses provide useful information for both the observer and the theorist.

For example, it is very important to note the exact times of contact between the edge of the Moon and the Sun so that they may be compared with those calculated from theory. Although eclipses can be predicted with considerable accuracy, it is still necessary to check these predictions against the actual event, so that our mathematical theory of the Moon's motion may be constantly improved.

Furthermore, it was formerly only at eclipses that we could observe certain interesting phenomena occurring of the surface of the Sun: the immense flame-like spurts of gas called prominences, not to mention the faint corona,

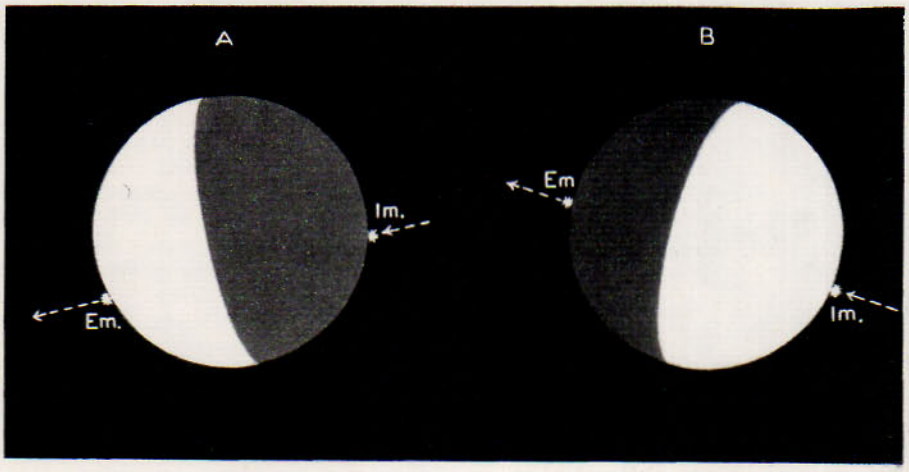


Fig. 329 — PROGRESS OF AN OCCULTATION NEAR LAST (A) AND FIRST (B) QUARTER

which is usually swamped by scattered light in the sky. Although it is now possible to observe these with special instruments, the conditions offered by an eclipse are much better and it is still worth while for astronomers to go off on an eclipse-watching expedition in some inaccessible part of the world, regardless of cost, discomfort and occasional dangers, and not infrequently, the risk of finding all their efforts wasted because of bad weather.

The actual information about the Sun's interior and immediate exterior which can be gathered in this way is dealt with in the chapter devoted to the Sun. One thing, however, may be noted: the time an astronomer has to observe an eclipse is very limited; since most total eclipses do not last longer than 3 minutes or so, and the maximum duration is only 7m 58s, as we saw earlier. Moreover, there are not more than about 50 total eclipses per century, of which a large number occur over inaccessible portions of the Earth's surface, where it is very difficult to install a temporary observatory with all the necessary equipment. Add to this the possibility that all these troubles will come to nought because of a single unlucky cloud passing in front of the Sun at the crucial moment. And even when the time is used to the full and when everything works out in the astronomer's favour, the total amount of time during which the eclipsed Sun can have been observed is no more than a few hours.

Fortunately, modern physics has found a partial remedy to overcome these disadvantages, and nowadays it is possible to study certain solar phenomena without waiting for a natural eclipse. It should not, however, be supposed that this renders the observation of eclipses superfluous, since the new methods do not as yet yield information about the whole of the Sun at the same time, nor do they permit observation of the outer parts of the corona.

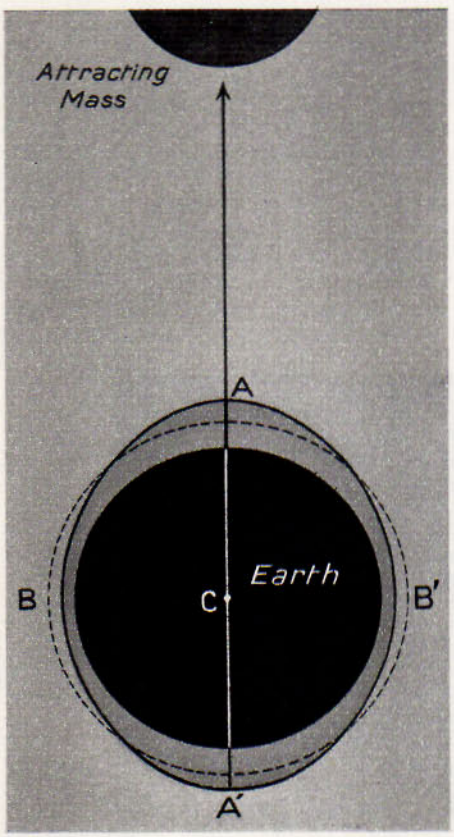


Fig. 330 — PRINCIPLE OF THE TIDES

OCCULTATIONS

As the Moon makes its monthly journey round the sky, it passes in front of any star or planet lying in its path in the same way that it passes in front of the Sun. When it eclipses a star or planet, however, the phenomenon is not called an eclipse but an occultation.

It will be remembered from earlier sections that the Moon does not travel along exactly the same path in the sky at every lunation; sometimes it is a little higher, sometimes a little lower. This means of course that it does not occult the same stars every month. After occulting a star one month, it will pass a little below or above it the next. This variation keeps within the Zodiacal belt, i.e. between the extreme northern and southern declinations that the Moon can reach. The question whether or not a planet will be occulted is even more complicated, because this depends not only on the Moon's motion but on the fact that the planet also is slowly moving round the sky.

As in the case of the solar eclipse, the factors determining from which part of the world an occultation will be seen and at what time are the relative sizes and positions of the three bodies in question: the star or planet, the Moon and the Earth. An occultation is visible only from certain parts of the Earth, and the time at which it is seen varies with the location. One place will see a small occultation whilst elsewhere the Moon will only make a very close approach (an appulse) and will therefore be in conjunction with the star or planet in question. Indeed, if the Earth had no atmosphere, this sort of approach would be visible every time there is a total eclipse of the Sun: we should be able to see the dark disc of the Moon silhouetted against the corona, and at every new moon it would be possible to see our satellite passing either above or below the Sun's disc.

In the case of a planetary occultation, we see the tiny disc of the planet disappearing behind the huge limb of the Moon much as we see the Sun setting below the horizon on Earth. Some time later the occultation terminates with the reappearance of the planet from behind the other edge of the Moon's disc. In contrast to this, stars, which are point sources of light, go out instantaneously as soon as the limb reaches them and reappear just as suddenly later. Therefore, the most useful observations of occultations are of these two contacts, respectively immersion and emersion.

Occultations have been used to try to resolve the problem of the lunar atmosphere. In the case of a planet, any air on the Moon would distort and diminish the brightness of the tiny disc, and some astronomers have reported observations supporting the presence of air from such appearances. At occultations of Jupiter and Saturn, a dark band has occasionally been reported along the edge of the Moon, visible against the bright disc of the planet. This, however, is almost certainly an effect of contrast along the line separating two surfaces of unequal brightness, the two planets being relatively dull compared with the Moon. This is confirmed by the fact that when a brighter planet is occulted — Mars or Venus, for example — no such effect is seen.

Occultations of stars, however, are even more valuable to the astronomer. Being to all intents and purposes instantaneous, a stellar occultation can be used not only to check for the presence of a lunar atmosphere but also as a way of measuring the Moon's exact position, which in turn gives us a check on the accuracy of our predictions. Actually, the apparent diameter of a star is a quite definite amount, although a star never shows a disc even in the most powerful telescopes; but by applying extremely sensitive electronic instruments, the time that

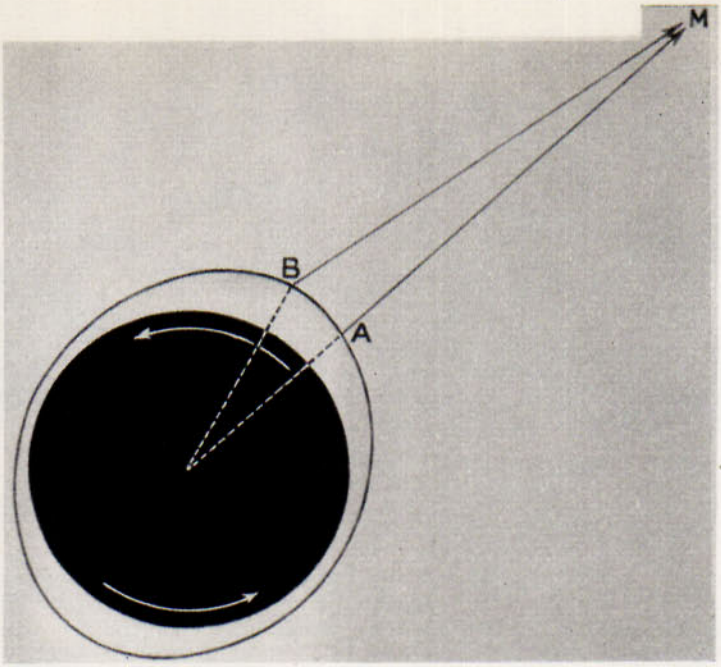


Fig. 331 — The rotating Earth tries to carry the tidal bulge around with it, from A to B, which explains the delay between the Moon's passage across the meridian (AM) and the arrival of high tide.

the Moon's edge takes to pass across the minute disc of a star can be measured, and knowing the angular velocity of the Moon, we can calculate the star's apparent diameter; from this the star's true diameter can be found if we know its distance (see Chapter 19).

Occultations may vary in appearance according to the phase of the Moon. During the first half of a lunation, the occultation takes place at the dark limb of the Moon, which may, however, be visible because of the Earth-light; during the second half of the lunation, conditions are exactly reversed. At full moon, both immersion and emersion take place at a fully illuminated edge of the disc.

THE TIDES

Among the most impressive phenomena we meet on the Earth are the tides, which are known even to people who do not see them in the normal course of their lives. This rhythmic movement of the sea, rising and falling

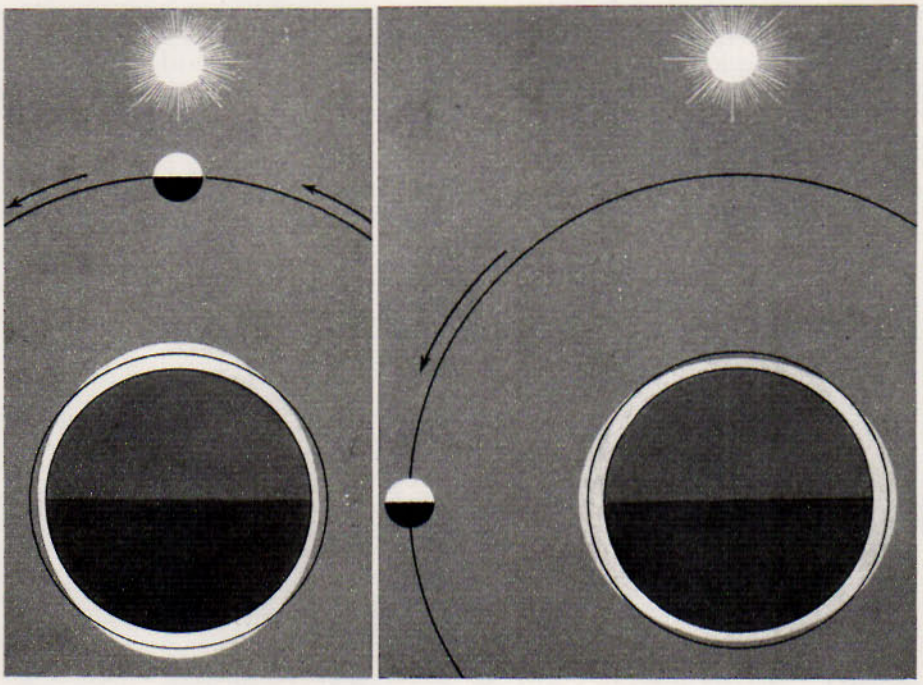


Fig. 332 — Spring tides are produced when the attraction of the Moon and that of the Sun combine; neap tides occur when they pull against each other.

Fig. 333 — A STEEP COASTLINE COMPOSED OF CLIFFS

Fig. 334 — A LOW, FLAT COASTLINE

every day, is due to the combined attraction of the Sun and the Moon.

Although the fundamental principles of tidal action are very simple, the way in which the tides actually behave is very complex, and depends not only on the astronomical causes just mentioned but also on the nature of the Earth's surface. Let us start by considering the astronomical factors.

Theory of the tides. — Newton was the first man to propound any satisfactory explanation of the tides, the subject forming part of his great work on universal gravitation.

Let us imagine the Earth to be perfectly spherical and covered with a uniform layer of water, at rest. If our planet were completely isolated in space, the layer of water would exactly match the spherical surface of the globe, but once some large body is introduced this figure of equilibrium is disturbed.

Under the attraction of this perturbing body, the water nearest to it will rise to form a bulge on that side of the Earth (A, Fig. 330), and since the volume of water available is finite, the water forming the bulge will have to be taken from other parts of the Earth, so that there will be a depression in the direction B-B′, at right angles to the direction from which the attraction is exerted. Moreover, in contrast to what might be expected, a second bulge will form at A′, opposite the first. The reason for this is simple: the force attracting the water will be stronger at A than at C, the centre of the Earth, since C is further away from the perturbing body than A. But in its turn, C is nearer than A′, and the weaker force at A′ causes the molecules or water there to lag behind the others. This results in the ovoid form (shown exaggerated in the diagram), the second bulge at A′ being slightly less pronounced than the first at A.

This is the theory behind the change in level of the oceans experienced as the tides.

The only two bodies large enough and near enough to have any perceptible tidal effect on the Earth are the Sun and the Moon. Although the Sun is by far the larger of the two, it is so far away that the Moon, with its much smaller mass at a much shorter distance, exerts the greater influence, since, as Newton showed, attraction decreases as the square of the distance. It is calculated that in spite of this inequality in both mass and distance, the attraction of the Moon is only about 2.2 times greater than that of the Sun. This means that although our satellite is the more important of the tide-raising forces, the Sun's influence is not negligible, and sometimes counteracts, sometimes supplements, the effects caused by the Moon.

The mechanism and amplitude of tides. — The alternation between high and low water which we know as the tides occurs with a period dependent on the Moon's synodic revolution period. High (spring) tides occur at the syzygies, i.e. at new or full moon, when the Sun and Moon are in line, at which times their individual gravitational attractions combine to raise a specially high tide in this direction, the tidal bulge A-A′ then being at its maximum height. Low (neap) tides occur at the half

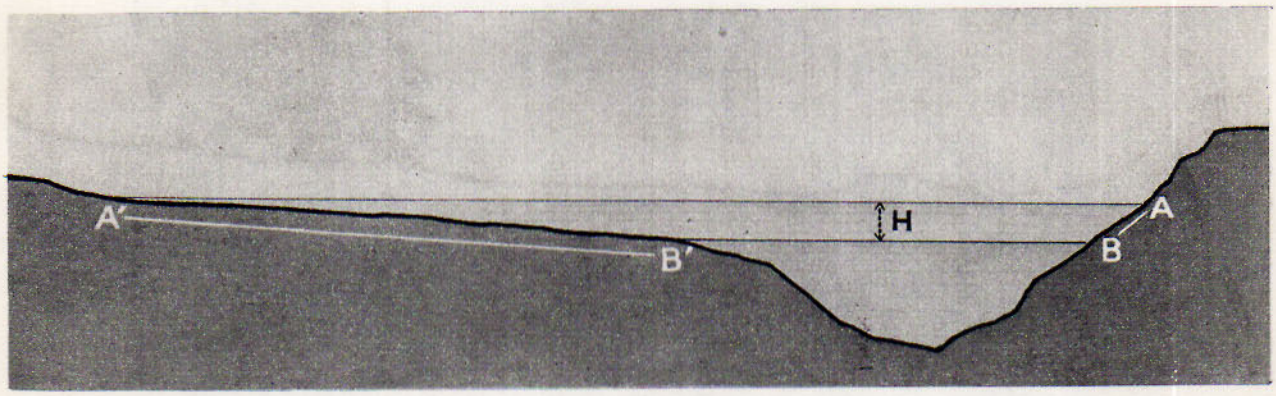


Fig. 335 — THE IMPORTANCE OF VARIATIONS IN THE LEVEL OF THE SEA ON A FLAT AND A STEEP COAST. If the level changes by the height H, the areas affected will be A—B on the steep coast, whereas on the flat coast it will be A′—B′, which is much the greater.

Fig. 336 — LINES ALONG THE SHORE SHOW THE HEIGHTS OF SUCCESSIVE HIGH TIDES

Fig. 337 — THE TIDE COMING UP A RIVER: THE CELEBRATED SEVERN BORE AT STONEBENCH, GLOUCESTERSHIRE. *(PA-Reuter photo)*

moon phase, when the Moon is at right angles to the Sun and their respective pulls operate against each other. The Moon pulls from one side and the Sun from another, which tends to even out the differences in the tidal level.

The relationship between the two extreme values (shown in Fig. 331) is simple. The ratio between the Moon's attraction and the Sun's is 2.2 to 1; at the syzygies they combine, adding up to 3.2 and forming the spring tides, while at the quadratures they subtract, giving 1.2 to produce the neap tides.

This relationship, however, is only a general one and can vary; in this connection we should recall what has been said about eclipses and the Moon's motion. The height of the tide is determined by the relative declinations of the Sun and Moon, and their distance from the Earth, being greatest when they are nearest the Earth and the equator. Therefore the highest tides of the year occur at the syzygies nearest the equinoxes, and the lowest at the quadratures nearest the solstices.

The above explanation applies only in connection with the relative positions of the Earth and Moon, but to understand the daily behaviour of the tides, it is necessary to allow for the Earth's axial rotation (Fig. 331).

The two main bulges still point towards the Moon, but since the Earth is rotating, it means that the surface must pass continuously beneath the stationary bulges, which give the high tides, and the intervening depressions, which give the low ones.

Thus the tides seem to move with respect to the continents, the alternations of high and low water matching the Moon's appearance in the sky. Our satellite moves in the same direction as the rotation of the Earth, but more slowly, crossing the meridian 50 minutes later every day, and so the interval between two similar phases of the tide is 24h 50m. Since there are two tidal bulges, there will be high tides at intervals of 12h 25m, with low tides between. The low tides, however, do not occur exactly half way between the two high ones, since the sea takes longer to ebb than to flow; this time varies from place to place.

It must be emphasised that the above account is very elementary and is designed only to explain the basic causes of tides. Their actual behaviour is much more complex and variable, and lies not in the province of astronomy but in that of hydrography and physical geography.

Careful observation of tidal behaviour at any particular place often leads to the conclusion that it is at complete variance with theory. At one place the difference between high and low water may be considerable, while a few miles away it is much less or even absent. There is a famous example of this on the French Channel coast: in the Mont St Michel Bay near Granville, the tidal range is 48 feet 5 inches, while at Cherbourg, a few miles up the coast, it is only 21 feet 10 inches. There are some even more startling differences in other parts of the world. In the Bay of Fundy (Canada) the maximum range is 65 feet 4 inches, at Porto-Gallegos (Argentina) it is about 58 feet, and at Portishead, on the Bristol Channel (England), 52 feet. On the other hand, in mid-Pacific the tidal range is very low (5 feet 4 inches in the Marquesas) and even lower in the Mediterranean. At Pola on the shores of the Adriatic it is only 4 feet 4 inches.

Positions having no tidal range at all are found in the centre of 'amphidromic systems', one of which exists in the North Sea between the coasts of Norfolk and of Holland. Tides oscillate about this point, giving high water on the Norfolk coast when it is low water on the Dutch coast, and vice versa. The times of high water proceed with this oscillation clockwise around the amphidromic point.

The range of these variations seems quite considerable when compared with the figure calculated for the height of the tidal bulge based on gravitational attraction, which is only 2 feet. The differences are due to the physical nature of the coastline. The tidal bulge moves rather as an undulation of the surface of the oceans, but its effects on the deep waters are transmitted along the coasts to masses of liquid which are smaller in volume because of the rise of the ocean bed. In certain bays, inlets and closed areas of the oceans a sort of pressure develops, and causes these tremendously high tidal ranges.

In ports and harbours mean high water (the average of all the high tides occurring there) may be expressed as a height above the mean tide level. As the pattern of high water heights is progressive between springs and neaps on each side of the mean high water height, a coefficient can be applied for any day, depending on the position of the day in a lunation and the solstices. This method, however, is very approximate, as is that of determining the time of high water by the number of days elapsed since the time of high water at full or new moon. Methods such as these have been replaced by new developments in tidal predictions, much of which is done by Tidal Predicting Machines. Predictions for many parts of the world are carried out by the Liverpool Tidal Institute, whose predictions of high and low water heights for Southend in 1957, when averaged, differed by as little as 1.8 inches when compared with the average recorded automatically at the site. At London Bridge the discrepancy was even smaller, being only 1.5 inches.

When it is considered that the complexities of the tides embrace many variables, such as astronomical position, wind and atmospheric pressure, it is surprising to note that in practice the theory of prediction goes hand in hand with observed levels when taken over a sufficiently long period. It is found, for instance, that an astronomical cycle of approximately 19 years is identifiable in some cases as a matter of inches in actual recordings, provided that the readings are averaged in 19-year blocks.

As well as causing the discrepancy between theoretical and actual tidal ranges, which was referred to above, the configuration of the coastline is responsible for determining the time between the Moon's meridian passage and the arrival of high tide at a particular place. This delay, which is practically constant, is called the establishment of the port. It varies considerably at different places; at Dunkirk it is 12h 13m, while at Lorient it is only 3h 32m.

The shape of the land-masses is another cause of the peculiar behaviour of the tides in the Pacific Ocean and the China Sea. In the Gulf of Tonkin, for example, there is only one high tide a day. This is just one instance of many peculiarities of behaviour, but to discuss the question fully is outside the scope of this work.

Before leaving the subject, however, we must make some mention of the slight tidal activity which occurs in the Earth's crust.

The explanations already given apply equally well to these crustal distortions. In their case, however, the variations are not very large between one place and another — unlike the ocean tides, whose behaviour is affected by every little feature of the coastline. Because of its elasticity, it is the whole of the crust which is deformed by the Moon's pull, the effect of which is smooth and uniform.

There is nothing obvious to show that these crustal tides do in fact occur, any more than there would be anything to reveal marine tides if we were floating on an ocean which covered the whole Earth. To detect any variation in height, the level of the sea must be measured against a coast which can be thought of as solidly fixed. In fact, however, the Earth's crust itself undergoes the

Fig. 338 — MONT ST MICHEL BAY (FRANCE) AT HIGH TIDE

Fig. 339 — MONT ST MICHEL BAY AT LOW TIDE

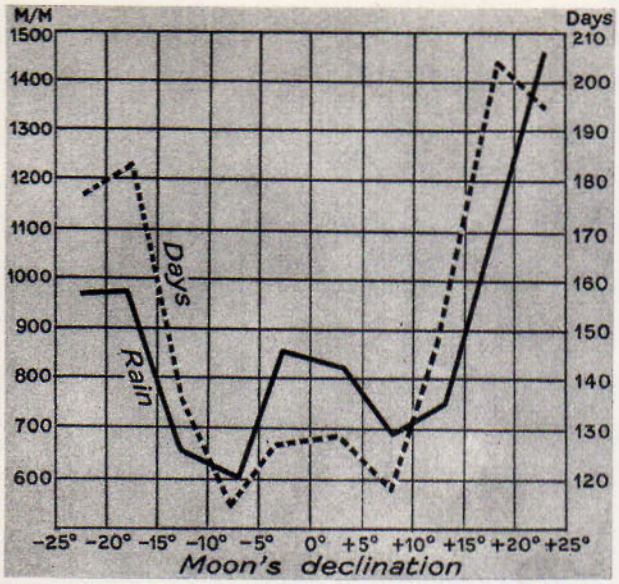


Fig. 340 — THE MOON'S INFLUENCE ON RAINFALL. Variations in the number of wet days per year and the total amount of rain as a function of the Moon's declination. After P. Luis Rodès (Èbre Observatory).

same sort of tidal action as the sea, and the height of the solid tide must be subtracted from the height of the liquid one. From this we find that the latter attains only two-thirds the theoretical value it ought to have if the terrestrial globe were truly rigid.

There is one last tidal effect which should be discussed. This is the 'bore' which occurs when a very high tide rushes up the estuary of a river against the normal flow, making a wave of water.

The explanation of this is simple. The flowing tide undergoes friction against the shallow bed of a river, and this retards the propagation of the wave-front, thus forming a sort of barrier behind which the rest of the tide begins to mount. When its pressure and height are sufficient, it begins to move fairly rapidly upstream as a wave breaking across the river, and the impressive noise of its passing can often be heard quite some distance away.

The occurrence of a bore depends in the main on the height of the tide, and is most likely when the tidal coefficient is greater than unity. But the nature of a river or estuary, as well as the state of the banks and mudflats, which are always shifting, must also be allowed for, as indeed must the direction and force of the wind and the barometric pressure. Most of these conditions cannot be known exactly in advance, and it is therefore impossible to predict the height and time of a bore with the same certainty as in the case of a tide. All one can say is that a bore usually occurs at a very high tide at the moment when the current reverses.

In the Seine estuary, the bore is particularly remarkable, since it has been known to reach a height of 15 feet and travel at 30 feet per second. In England the most famous example is the Severn Bore, which reaches an unusually long distance inland (Fig. 337).

The action of the tides. — The main action of the tides is to exert a frictional force against the Earth. Since the tidal bulge always attempts to follow the Moon, with the Earth rotating beneath it, the constant friction thus set up between the water and the ocean beds and coastlines tends to act as a brake on the Earth's rotation. Although very small, it is enough to produce a continuous slowing down of the Earth and a consequent lengthening of the sidereal day. This slowing down does not appear to be perfectly regular, and this is probably due to various internal causes. These no doubt are in some way connected with external astronomical forces.

The fact that the Moon always turns the same face towards the Earth is the result of some earlier tidal action under the influence of the Earth when our satellite was still plastic. It is possible to say, therefore, that by this mutual interaction the Earth-Moon system is tending towards a state of equilibrium with both bodies revolving about their common centre of gravity, and each keeping the same face turned towards each other. This, however, will take place only in the very distant future, many millions of years hence.

Of more immediate interest to us is the continuous work of erosion carried out by the tides round the edges of the terrestrial land-masses. This process forms part of the eternal cycle of continent-building and levelling.

In their attack on our coastlines, the waves are greatly helped by the ebb and flow of the tide, particularly in places where the tidal range is considerable. It may be pointed out that when there is a storm under certain conditions, coinciding with one of the particularly high tides occurring near the equinoxes, the general level of the sea may rise as much as 3 feet or more above the height of the tide.

The force needed to move the ocean waters in this way is tremendous, and many schemes have been proposed to harness it for the benefit of man. The principle is to allow the water to rise in large areas contained behind dams. When the tide turns, the sluices are closed and the trapped water can then be fed through turbines to generate electricity. It is enough to watch the tide to see how powerful the force is, but figures give a more exact idea. Let us consider the famous Mont St Michel Bay in France: every high tide involves the shifting of a *thousand million tons* of water, and this is repeated twice a day. If properly used this would yield 3 million horsepower, which is greater than the whole of Niagara Falls.

From our point of view, all this energy is going to waste, and is frequently working against us by wearing away our coasts. But when Man learns to tame these forces and put them to good use, he will find in the Moon a new and valued friend, the source of abundant and inexhaustible energy, aptly known by the name of 'blue coal'.

The influence of the Moon. — In addition to the purely mechanical activity of the tides, the Moon has been blamed for all sorts of imagined influences on the Earth and its inhabitants. Indeed there is enough material here for a whole book to itself, and we shall therefore have to confine ourselves to the most important points.

The Moon's influence on our atmosphere is implicitly believed in by most people. The weather is held to be regulated by the four main phases of the Moon, the new moon being considered particularly significant.

Many people claim to have proved such correlations and have cited statistical evidence in favour. This has led them to formulate more explicit laws than those contained in the old superstitions and proverbs. If these laws are true they should allow accurate weather forecasts to be made.

The best way to test such claims is obviously to concentrate on the study of one particular phase of the Moon, since it will bring out any lunar action in the atmosphere analogous to tides in the ocean. We can then consider how atmospheric tides show themselves in the vicissitudes of the weather. Let us not, however, be too

eager to accept explanations of this sort. Although such tides do certainly exist theoretically, producing slight changes in barometric pressure, they are completely disguised by barometric changes due to the constant succession of weather systems passing across the face of the planet. Effects of lunar action would be detectable only on rare occasions. At such times, the Moon's influence might just tip the balance one way or the other to produce some discernible effect. It may be remarked that the observational evidence brought forward is very localised, and it is well known what a large part local geography plays in weather conditions. Consequently it may well happen that in one place a predicted change does take place, whilst nothing of the sort occurs a few miles away.

Until further evidence is forthcoming, we must keep an open mind on the question of changes due solely to the influence of the Moon, since the percentage of predictions which come true is not high enough to remove all doubt. In fact many statistical investigations into the matter have given negative results. The fact that meteorologists manage to forecast the weather quite accurately without taking the Moon's influence into account may be taken as sufficient indication.

These conclusions were reached by the late P. Luis Rodès, Director of the Ebro Observatory, Spain. Observations collected over a period of twenty years show that there tended to be a greater amount of rain when the Moon was at perigee than at apogee. Similar increases in rainfall were noted when the Moon was at its greatest declination from the equator, particularly to the north (Fig. 340) and also at the time of upper and lower meridian passage. Further influence was seen in the decrease in wind speed and the increase in cloud at the time of the full moon. No appreciable meteorological variations were noted at the time of the new moon, which would seem to weaken the case for the old superstition.

This evidence is highly interesting but it must be borne in mind that the region studied was very limited and that the variations mentioned were more in the nature of general tendencies than clearly marked changes. There is no doubt that more observations are urgently needed, preferably on a world-wide scale and over long periods of time, to settle the question whether the Moon has any influence on the world's weather.

A common belief is that the Moon 'eats up the clouds'. Many people attempt to rationalise this belief by appealing to the Moon's radiation of heat received from the Sun. Unfortunately it requires extremely delicate scientific instruments even to detect this heat and there is certainly nothing like enough to evaporate clouds. Against this it is said that this heat is absorbed in the upper reaches of the atmosphere where things are supposed to be different. On the other hand, it must be remembered how unstable the conditions are under which water vapour may be suspended in the atmosphere, so that the slightest disturbance may upset the delicate equilibrium. Thus condensation products can pass from the visible to the invisible state, and vice versa, with great rapidity. The proof of this instability is found in the behaviour of very small clouds during the daytime: they form and disappear in a matter of minutes, and this is probably the result of atmospheric movements involving slight changes in altitude. There is no reason why this should not happen equally at night, in which case the part played by the Moon is simply to make the occurrence visible. In spite of this, it may still be claimed that the Moon plays a more active rôle — rather like the last drop of water which makes the jug overflow. This, however, is pure speculation, and although — because it is difficult to distinguish pure coincidence from cause and effect — we cannot say for certain, the evidence seems to be to the contrary.

The Moon's radiation is blamed for more ill effects of this sort, for example the famous 'April Moon', feared by market gardeners and horticulturalists generally.

The 'April Moon' is the name given to the lunation beginning in April, with the full moon in either April or May. According to popular superstition, our satellite has a disastrous effect on buds and young flowers exposed to it on a clear night. They turn a rusty brown colour as if attacked by frost, although the thermometer temperature does not fall low enough for this. The best 'proof' of the Moon's guilt in this matter is that nothing similar occurs when the sky is cloudy. In short the Moon's rays are credited with the property of refrigeration! But if they have this property, why only at this particular time of year? Moreover, this contradicts the suggestion that it is the Moon's heat which disperses clouds. The superstitious cannot have it both ways . . .

Let us quickly exculpate our satellite: it is merely the witness, not the cause of the ill effects in question. The true explanation is a purely physical one.

The surface of the ground, and anything on it, including vegetation, can lose heat at night by radiation into space. This brings its temperature down below that of the ambient air. The reason this occurs only when the weather is dry and clear is simply that clouds act as a sort of blanket and keep in the heat. Consequently, when the atmosphere is calm and clear, the ground and any plants on it may be in the grip of frost, even quite seriously so, whilst a thermometer in the surrounding air may still register several degrees above freezing point. During the spring months the right conditions are often present, for at this time of year the amount of heat stored up by the ground is still small, whilst clear nights become increasingly frequent. Thus the season is characterised by the freezing of buds and blossom, but as the thermometer does not give any warning of frost and since the Moon is often present, it was a natural but an erroneous step to blame the 'Queen of Night'. It may be noted that this nocturnal deposition of frost still occurs even when the Moon is not in the sky.

The Moon is often credited by the superstitious with the power to influence the growth of trees and the

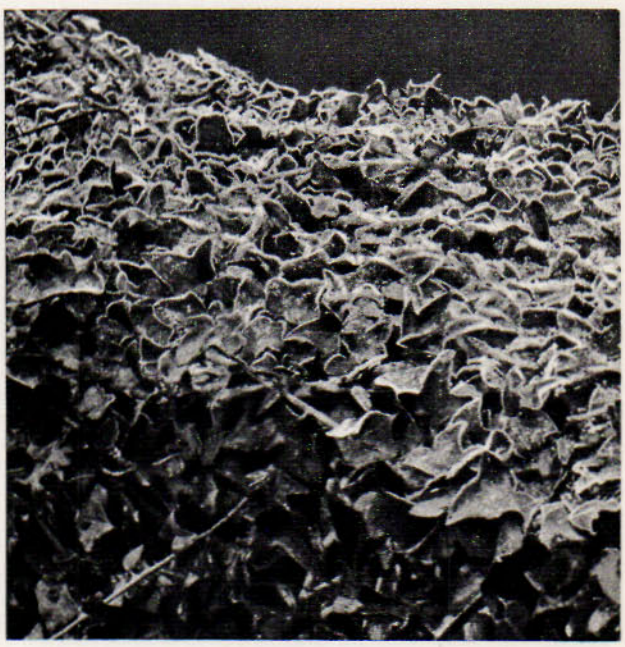


Fig. 341 — WHITE HOARFROST APPEARS ON OBJECTS LOSING HEAT INTO SPACE AT NIGHT

germination of seeds, etc. Human beings too, especially the more unstable type, are said to be affected, as is the egg-laying and even the whole reproductive cycle of some animals. In this context certain observations are worthy of attention. They concern the extremely peculiar behaviour of certain marine worms, belonging to the family of Annelida. Although they never show themselves as a rule, in October and November and only at the last quarter of the Moon they can be seen in tremendous numbers on the surface of the sea, performing a remarkable 'nuptial dance'. Is this merely a striking coincidence or is it a case of cause and effect?

The large and interesting questions which the above observations raise are beyond the scope of this book; they are, indeed, difficult ones to answer and the first task is to disentangle preconceived ideas and false interpretations from the true facts, after which it is necessary to undertake a long and strictly controlled series of experiments and other researches to determine the exact causes.

In the meantime, all we can do is to refuse to believe in the many effects attributed to the Moon without sufficient evidence. It may be added that not even those which seem most clearly attested can be taken as the basis of useful prognostics, being slight, doubtful and lacking in general application.

These last few pages have taken us a little way from the subject of astronomy properly speaking. But after making an astronomical acquaintance with the Moon, it did not seem inappropriate to recall the more immediate importance of our satellite for the inhabitants of the Earth.

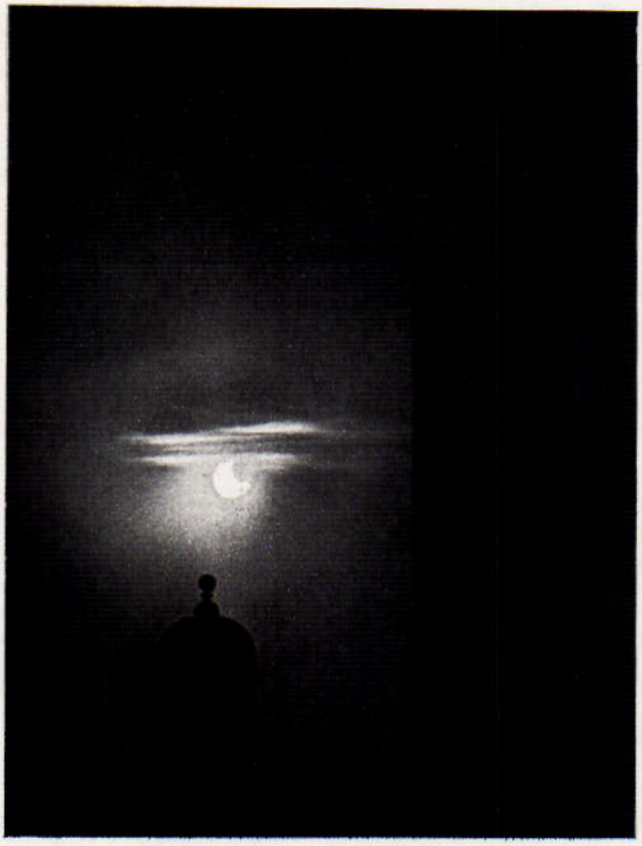


Fig. 342 — PARTIAL ECLIPSE OF THE SUN, 19 APRIL 1939, AT PARIS. *(Photo, France-Presse)*

Fig. 343 — A CONJUNCTION BETWEEN PLANETS AND THE MOON IS A SPLENDID SIGHT IN THE TWILIGHT SKY

CHAPTER 7

THE PLANETS

Our visit to the Moon did not take us far from home, although it was far enough by all ordinary standards. Now we must make the much larger jump through space to reach the planets.

The nature of the planets and of the system they form around the Sun was discussed in an earlier chapter. Now let us examine each of them individually, studying in greater detail their movements, composition and physical characteristics generally. In this way it will be possible to judge them as 'worlds', thus answering a question which nearly everyone asks at some time: how closely do they resemble the Earth?

At first sight it might seem that this question could easily be answered nowadays by virtue of our refined modern methods, but this is true only in certain respects. For example, the real diameters of the planets have been measured exactly and the planets have even been weighed — mathematically. From this certain comparisons can be drawn between them and our own world. Furthermore, we know how they move round the Sun and rotate upon their axes, which gives us the length of their night and day, their year and the progress of the seasons. But detailed information about the exact composition of these planets, and the various phenomena occuring on and within them, is not available. Much of our knowledge is derived simply from observations of their surfaces; this, as we have seen, is a very delicate job, fraught with difficulties of interpretation due both to the distances involved and to other causes which we shall discuss later. Fortunately, in recent years the application of astrophysical methods to planetary observation has yielded a great deal of accurate information to fill gaps in the knowledge gained from visual work alone. For the moment, however, let us confine ourselves to a few general remarks and make a comparison between the distant planets and the Moon, the nearest heavenly body and the one that we see most clearly.

In our earlier discussion of the Moon, it was shown that the limits of lunar observation by traditional methods have now been reached, in so far as no increase in the size of our instruments will give a corresponding increase in the minuteness of the detail seen on the Moon's surface; the reason for this is that disturbances in the Earth's atmosphere frequently introduce distortions into the image which are much larger than the theoretical resolving power of the instrument. This means that it is becoming increasingly unprofitable to employ larger and larger telescopes for lunar work, since under average conditions of observation smaller instruments will do the job just as well.

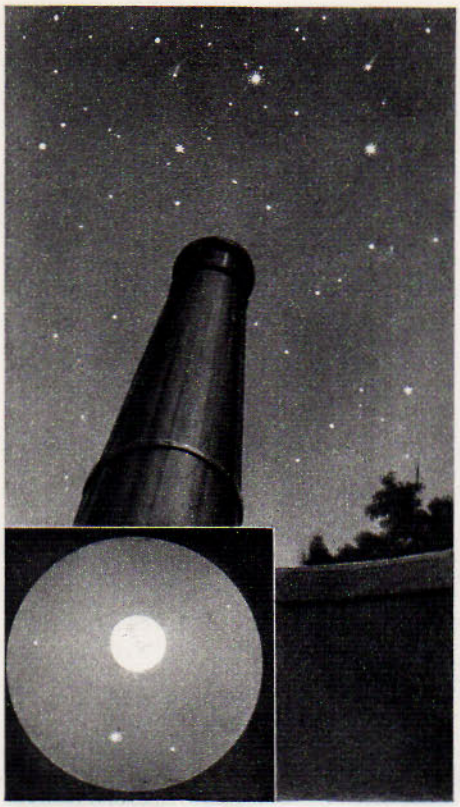


Fig. 344 — WITH THE NAKED EYE MARS LOOKS LIKE A BRIGHT STAR; IN THE TELESCOPE IT SHOWS A SIZEABLE DISC

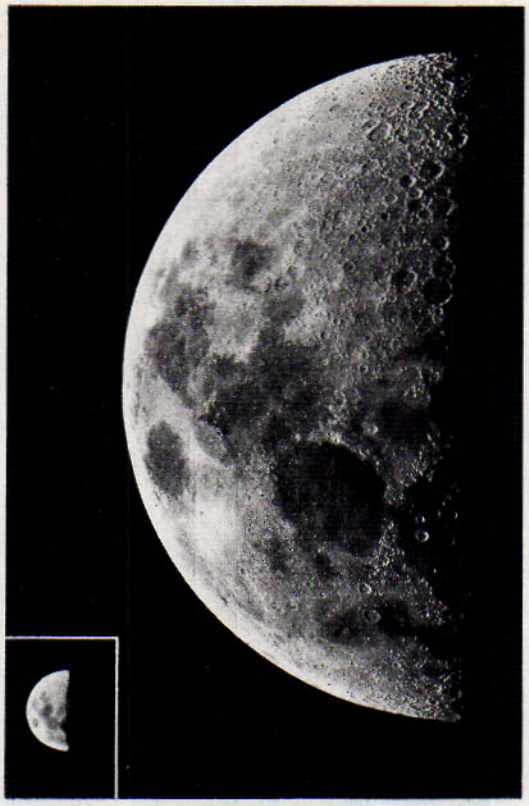


Fig. 345 — THESE PICTURES OF THE MOON ON DIFFERENT SCALES SHOW HOW FINE DETAILS MERGE INTO LARGE UNIFORM AREAS ON PLANETARY DISCS

These difficulties are much more acute in observations of the planets. Under the most favourable conditions, i.e. when they are in opposition with the Sun and fully illuminated, even the nearest planets are more than 150 times as far away as the Moon, and the others are correspondingly further. No more need be said to show that we cannot hope to distinguish structural detail on the surface of a planet in the same way as we can on the Moon.

As an example, let us consider some feature on the Moon which, under a given telescopic power, appears as a spot, almost a point. At the Moon's distance, this spot will correspond to, say, an object a mile across. Now the difficulty, of course, is to know whether that little spot in the telescope actually represents one distinct formation, or a series of formations too small to be seen separately but visible when merged together by insufficient resolving power, coupled with atmospheric blurring.

And so when we observe the planets through the telescope, we only see the broader features in their general outlines, but we know that these features must be made up of a whole series of smaller details of varying shapes which are too small to be seen individually.

The importance of this is easily understood. The Moon can, it is true, be examined at close quarters and many questions have been answered satisfactorily in this way. Although it is impossible to make out the finest structural details, the general lay of the rocks of which the Moon is formed, as well as relatively small surface features, can be observed through the telescope, and this does provide valuable information to be incorporated in theories about the Moon. For the planets, however, the situation is much more difficult since even on the nearest of them the smallest object that can be seen is between 50 and 100 miles across; even then it appears only as a minute spot, the finest details of its structure lying beyond the resolving power at our command.

To sum up then, the details of various surface features, their patterns of shade and colour, which make up the familiar appearance of each planet, actually represent enormous areas of ground and can be used only as clues from which to deduce the real nature of the surface. It goes without saying that the further away the planet, the greater the difficulties of interpretation. Consider, for example, the dark patches found along the system of belts parallel to the equator of Jupiter. Jupiter is much more distant than Mars. In this case, therefore, these small patches are not merely 100 miles across but correspond to features that are comparable in size with the Earth itself. We know that they must be composed of finer structural elements, but we have no idea as to their real appearance.

Nevertheless, our ideas have become more certain as astronomical techniques have improved. Thus the first lunar observers, using a naked eye or insufficient optical power, thought that the large grey patches scattered over the Moon's surface were actually oceans, because when seen in this way they appear to be uniform in nature. Greater optical power, however, revealed structural details on the surface of these 'seas', and it became obvious that the earlier notions were incorrect, that these large plains, in addition to having many small features on them, were completely dry.

The planets are always seen as quite small discs, and even those with the largest apparent diameter are about 40 times smaller than the Moon. Consequently, quite a powerful instrument is required to show a planet as well as we can see the Moon with the naked eye or with

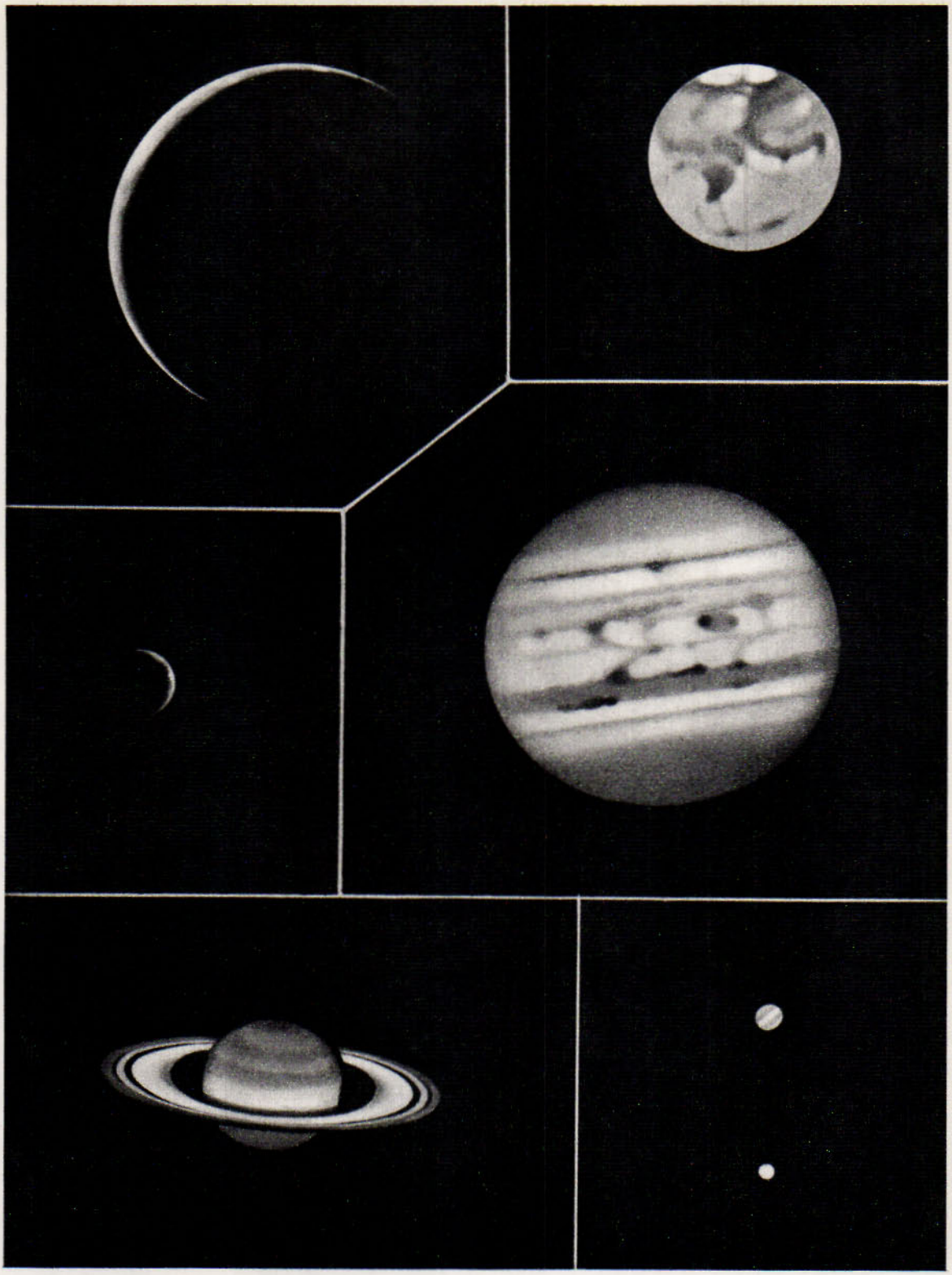


Fig. 346 — COMPARISON OF THE APPARENT SIZES OF THE DIFFERENT PLANETS AS SEEN THROUGH A TELESCOPE. *Top left to bottom right:* VENUS, MARS, MERCURY, JUPITER, SATURN, URANUS, NEPTUNE. (Pluto is not shown.)

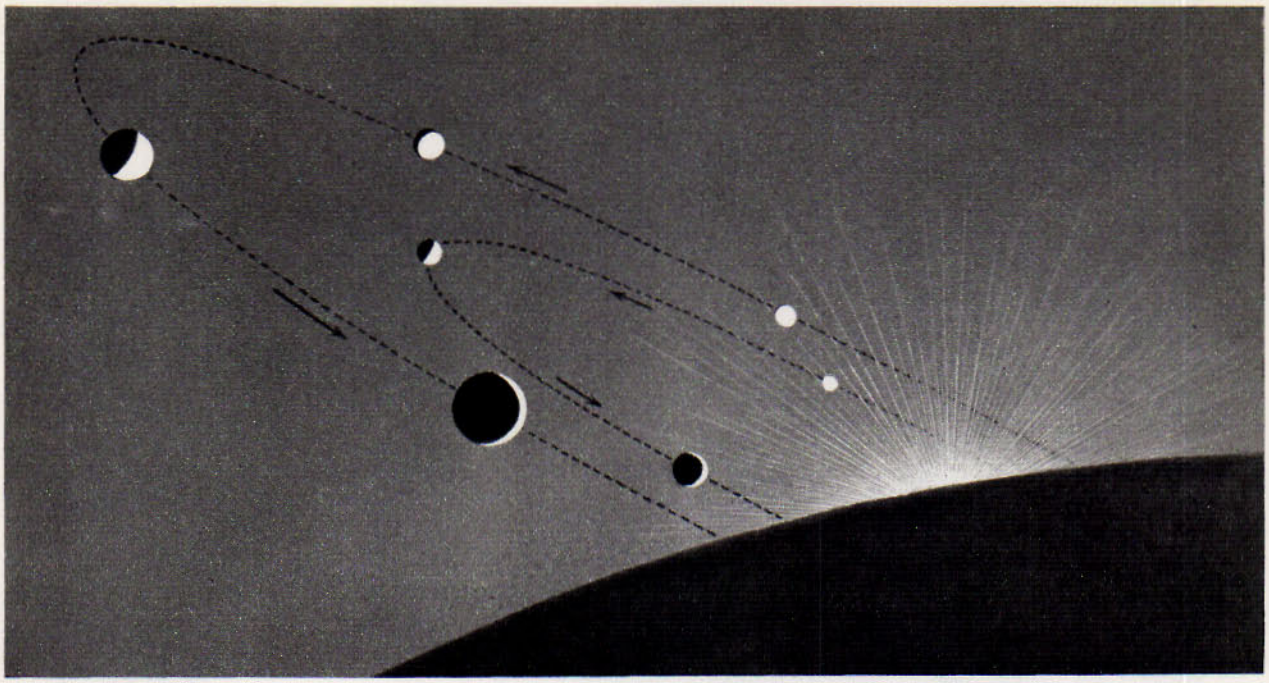


Fig. 347 — THE APPEARANCE OF MERCURY AND VENUS, WHICH GO ROUND THE SUN INSIDE THE ORBIT OF THE EARTH

a very low magnifying power: in such conditions the characteristic details of the Moon are not visible as such, but take on the appearance of darkish patches on the surface of the disc. It is therefore hardly surprising that differences of opinion arise about markings observed on small planetary dics, markings which are, as we have seen, merely the general impression left by a whole series of faint, smaller features on the threshold of vision. Moreover, the observational problem is frequently aggravated by atmospheric turbulence which spoils the telescopic image: this is the curse of planetary observers. Further complications are introduced by the relative positions of the bodies concerned — Earth, Sun and planet — which means that the conditions of illumination of a planetary disc are constantly changing, parts of certain planets being hidden for long periods at a time.

One point must be emphasised before proceeding further: one of the most important links in the chain of observation is the observer's visual acuity and skill, which determine how accurately he will interpret the light-signals falling on his retina. Drawings of the same planet made at the same time by different observers, all in equally good faith, are frequently so unlike as to make them very difficult to compare. However, the patient accumulation of repeated observations usually leads to new techniques, and in turn these give further information which may be checked by comparison with knowledge gained on Earth, so that we now have a body of reasonably reliable knowledge about each different planet. This will be the substance of the following sections.

Orbit and motion of Mercury. — This, the nearest planet to the Sun, travels along a comparatively eccentric orbit. Its mean distance is 36,000,000 miles; its perihelion distance is only 28,750,000 miles whilst its aphelion distance is 43,750,000 miles. Consequently its velocity varies considerably at various points round its orbit, of which it makes one complete circuit in 87.97 days. Mercury's velocity is 35.5 miles per second at perihelion and 24.375 miles per second at aphelion. This inequality has some important consequences, which will be discussed later.

In our earlier discussion of the Solar System, we studied the behaviour of a planet situated inside the orbit of the Earth, and we may recall that its apparent movements consist of a periodic oscillation from one side of the Sun to the other. In the case of Mercury the greatest distance reached to west or east of the Sun (called elongation) undergoes considerable variation because of the planet's

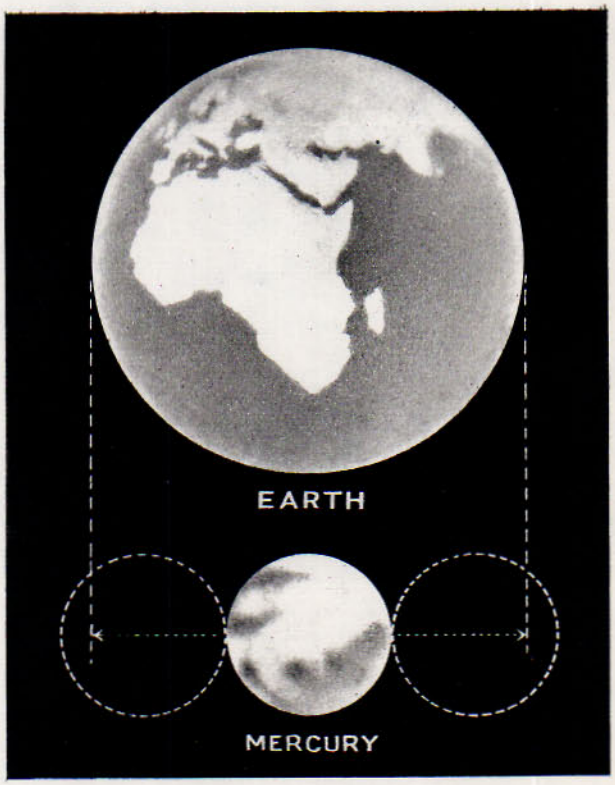


Fig. 348 — THE SIZE OF MERCURY COMPARED WITH THE EARTH

Fig. 349 — ALWAYS CLOSELY PRECEDING OR FOLLOWING THE SUN, MERCURY SHINES BRIGHTLY IN THE TWILIGHT AT GREATEST ELONGATION

orbital eccentricity. The relative positions of the three bodies at different times are shown in Fig. 350, from which it appears that the angle between Mercury and the Sun at greatest elongation attains a maximum of 27° 5′, but at some elongations does not exceed 17° 50′. In any case, even the greatest possible angle is rather small, which means that Mercury always appears to stay close to the Sun in the sky, following or preceding it by an interval of 2h 15m at the most. Consequently, Mercury is visible only in the evening and morning twilight, low down near the horizon, and it will be remembered from earlier remarks about the behaviour of the atmosphere that these conditions are not conducive to satisfactory telescopic observation, to say the least of it. In order to be of any use, observations of Mercury must be made when the planet is high in the sky, but then, unfortunately, the Sun fills the atmosphere with intense scattered light, and the planet's image, which can be inspected only with accurately directed instruments, is very difficult to observe.

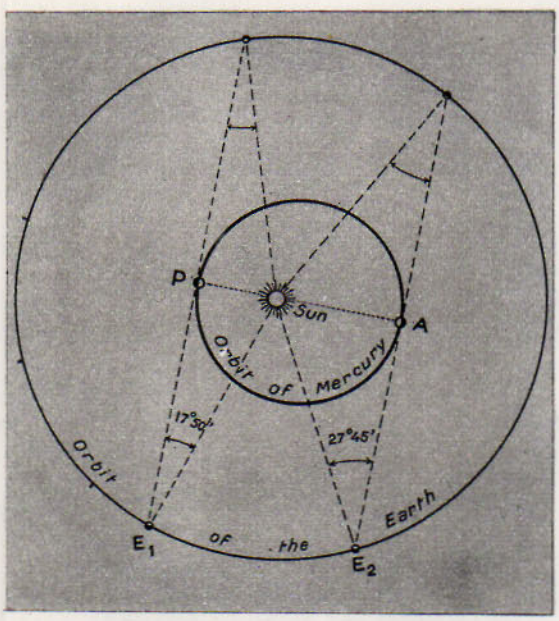


Fig. 350 — RELATIONSHIP BETWEEN THE ORBITS OF MERCURY AND THE EARTH

Further difficulties arise from the rapidity of Mercury's orbital motion, which confines its appearances to a few days at a time — the week or so either side of maximum elongation — and to be of any use this must coincide with a period of good, clear, still weather. This all means that any one person can observe the planet only at very infrequent intervals. It may be recalled in this connection that at the end of a very long and industrious career, Copernicus himself is reputed to have complained of never having had a chance to see Mercury through the misty skies of his native Poland.

Let us now examine the problem of Mercury's orbital motion in more detail. In the first place, since the Earth is moving in the same direction but less rapidly than Mercury, the two bodies return to the same relative positions after an interval of 116 days. During this time, Mercury will pass through the following series of phases: —

Beginning at greatest elongation east of the Sun, Mercury is well visible in the evening after sunset. Then the planet will apparently start to move with increasing speed towards the Sun. This, however, is the result of seeing its orbit in perspective, and Mercury is actually passing between the Earth and Sun, where it becomes lost in the glare of the latter. This is called inferior conjunction. Continuing in the same direction, Mercury will emerge on the other side of the Sun into the morning sky, finally reaching greatest western elongation. Between these two extremes (eastern and western elongation) there is an interval of 44 days on the average. After the morning appearance, the planet begins to move back once more towards the Sun, passing round the far side of the latter at superior conjunction, then moving to the east into the evening sky, and back to greatest elongation. The relative movement of the Earth and Mercury means that the time between greatest western and greatest eastern elongation, when Mercury is travelling along the part of its orbit furthest from the Earth, is 72 days on the average.

In this game of hide and seek, as the 'winged messenger' rushes from one side of the Sun to the other, Mercury can normally be seen with the naked eye only at or near greatest elongation. It usually begins to come into view some 17 to 10 days before evening elongation and vanishes 6 to 7 days after. The same periods, in reverse

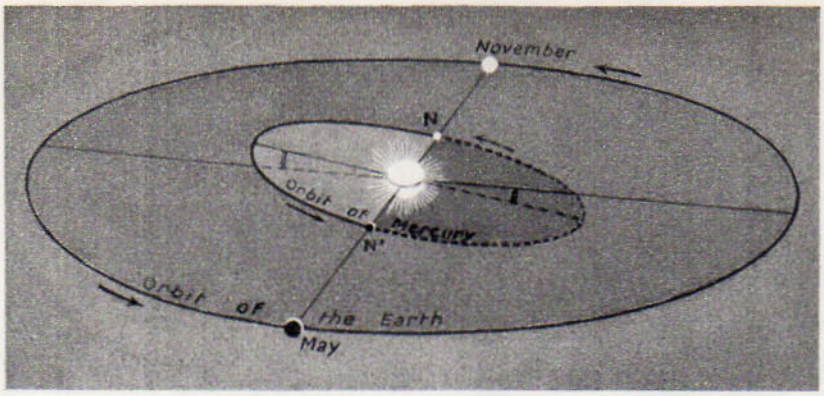


Fig. 351 — THE INCLINATION OF MERCURY'S ORBIT DETERMINES THE TIMES OF ITS TRANSITS ACROSS THE SUN

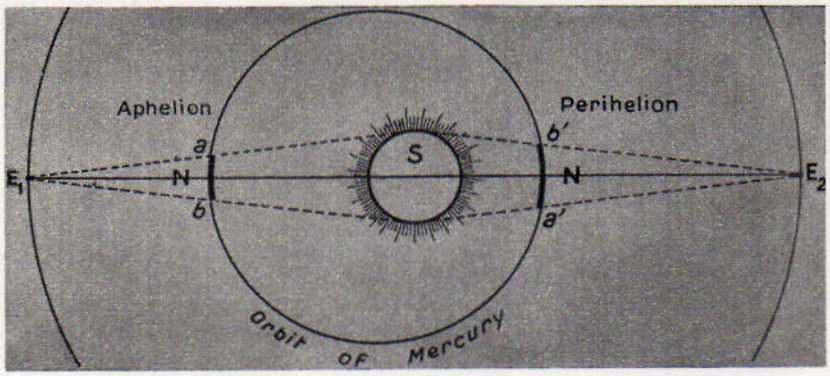


Fig. 352 — For the Earth at B_1 (May) Mercury, passing across the line of the nodes N, will be seen in transit while between *a* and *b* in front of the Sun S. For E_2 (November) the angle *a' b'* is larger and transits will therefore be more numerous.

order, can be applied to its appearance in the morning sky. Outside these brief intervals, Mercury may be seen in broad daylight through a telescope which has been pointed in exactly the right direction.

The general orbital motion considered up to now is complicated further by a number of things, including the marked orbital eccentricity and the angle of 7° which the orbit makes with the plane of the ecliptic. This means that when inferior conjunction coincides with the line of the nodes, Mercury can be seen as a black dot passing across the face of the Sun. These phenomena occur only at very infrequent intervals because of the way in which the movement of the Earth combines with that of Mercury.

Transits of Mercury occur at intervals of 13, 7, 10 and 3 years. The Earth arrives at the points where the line of the nodes of Mercury's orbit cuts the ecliptic on 7 May and 9 November. In the years when they do occur, transits take place either on or near these two dates. In May, at the descending node, Mercury is situated near aphelion, i.e. at its closest to the Earth, and its apparent diameter is then approximately 12″. The laws of perspective restrict the angular limits either side of the node within which Mercury can be seen silhouetted against the Sun. In November the opposite situation occurs. Mercury is further away and its apparent diameter is only 9·8″, but there is greater latitude either side of the node for transits to occur (Fig. 352). Consequently November transits are more frequent than May transits in a ratio of about 7 : 3. Their duration is likewise very variable, depending on the length of the chord traced by the planet across the Sun's disc, and its speed of transit. The maximum duration is nearly 9 hours, but it is sometimes no more than a few minutes if Mercury's path is almost tangential to the edge of the Sun's disc. And sometimes

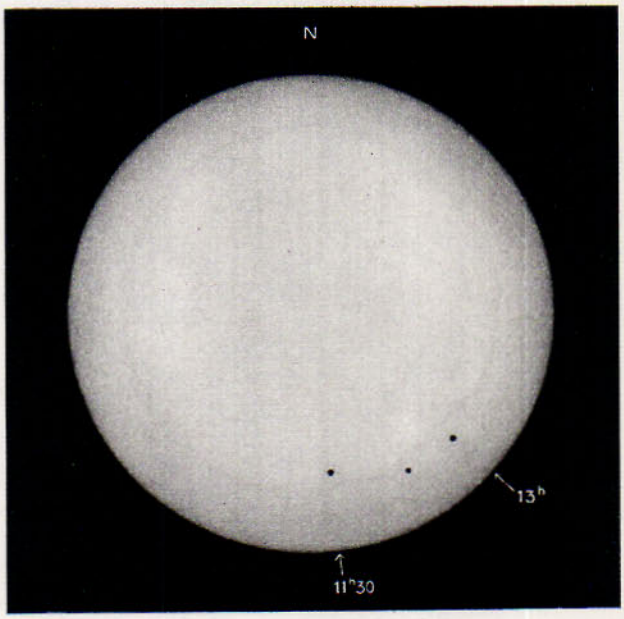


Fig. 353 — TRANSIT OF MERCURY, 6 NOVEMBER 1914

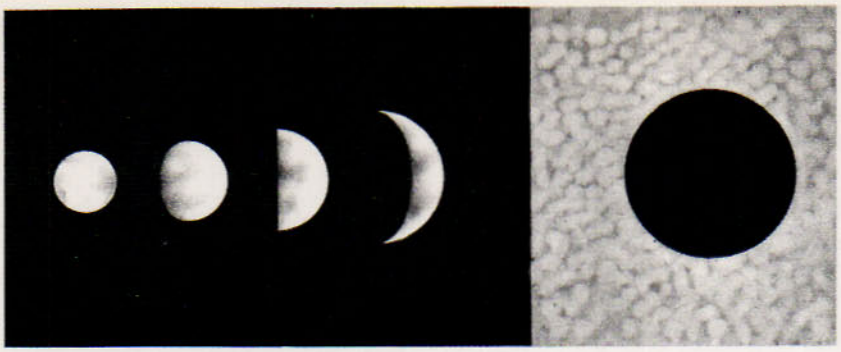


Fig. 354 — VARIATIONS IN MERCURY'S APPARENT DIAMETER ACCORDING TO ITS PHASE

Mercury only just touches the edge of the Sun without actually passing in front of it. There is then no transit, properly speaking.

The last three transits were on 12 November 1940, 12 November 1953 and 6 Nov 1960. The following are all that will take place before the end of the twentieth century:

1970 May 6
1973 Nov 11
1986 Nov 12
1999 Nov 14

It is only at these times that Mercury can be seen with its maximum apparent diameter. At all other positions in the sky, and therefore at all other phases, the planet is further away and smaller. It is at its minimum diameter when the phase is greatest (Fig. 354). Because of its small true diameter, Mercury's disc is always very reduced in size as compared with the other planets, and this, of course, increases the observational difficulties even further, in particular exaggerating the effects of disturbances in the Earth's atmosphere.

The rotation of Mercury. — Through a fairly large telescope, the surface of Mercury is seen to be covered with faint, hazy markings. From changes in the position of these markings, it was first thought that the planet's rotation period was comparable with the Earth's. At the beginning of the 19th century, Schröter estimated it as 24h 0m 50s, taking the axial inclination as 20° to the plane of the orbit.

This rotation period was accepted for many years until in 1891 Schiaparelli announced that instead of turning rapidly on its axis, Mercury makes only one complete rotation during its period of revolution round the Sun: in other words, its day equals its year. The planet consequently keeps the same face turned towards the Sun, just as the Moon does towards the Earth. In the case of the Moon, as we have seen, the eccentricity of the orbit introduces a 'balancing' movement called libration. The same thing occurs in Mercury's case, but is even more pronounced because of the latter's greater orbital eccentricity, so that the planet swings 23° 7′ in longitude alternately either side of its mean position. Consequently there is a gradual shift in the position of the surface markings with respect to the terminator, and it is this

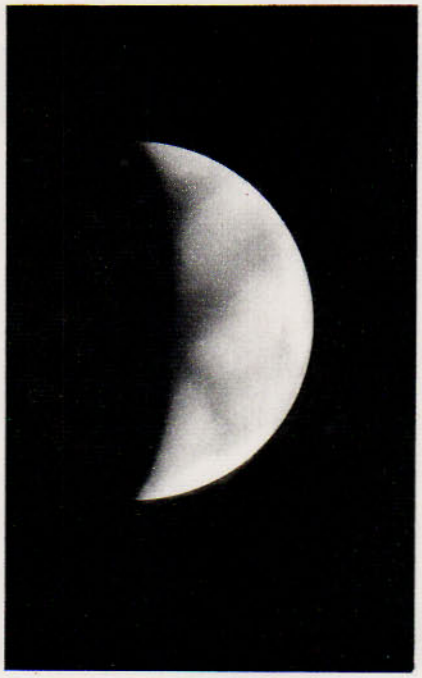


Fig. 355 — APPEARANCE OF MERCURY AFTER WESTERN ELONGATION

Fig. 356 — APPEARANCE OF MERCURY AFTER EASTERN ELONGATION

which erroneously suggested a rotation period of 24 hours, since practically the same view of the planet is obtained at the same time when observed in the morning or evening, after a certain interval.

This determination of Mercury's rotation period did not confirm the axial inclination suggested by Schröter; this was derived from the observation of certain surface markings which he considered to be on the planet's equator. Even now it is impossible to give a completely accurate figure, since exact observations are so difficult to make owing to the small size of the disc; but in seems likely that Mercury's axis is not quite perpendicular to the plane of the orbit, though the divergence cannot be very great.

The surface of Mercury. — Because of the way in which it moves round the Sun, only a little more than half of Mercury's globe is visible from the Earth, the other remaining in eternal night.

Whether in the crescent, half-moon or gibbous phase, Mercury always appears very small and this makes it a difficult object to observe in detail. Normal observational methods reveal only the position of the terminator and a few differences in the surface brightness at various places on the disc and along the terminator.

Generally speaking, Mercury has a quite pronounced yellowish colour, with fairly distinct grey patches here and there, whilst there are also some small areas of a lighter tone, verging on white. In practice these markings are rarely shown exactly alike by different observers for the reasons mentioned in the introduction, but it can be said with some certainty from their general distribution and always easily recognisable pattern that they are permanent features of the surface; this is also true of the various whitish regions.

Independent of this surface detail, there is a noticeable darkening along the terminator, where the disc fades off into a vague, ill defined blur. Moreover, there are certain irregularities which are always seen in the same circumstances: e. g. the truncated appearance of the southern horn of the crescent, which corresponds to a particular phase. There may be two reasons for this: either the nature of the ground is such that it rapidly decreases in brightness when the incident light strikes in below a certain angle, or — and this is more likely — there are real differences in level producing irregularities in the terminator in the same way as on the Moon. The proximity and size of the Moon show up these differences for what they are, but Mercury is so small and far away that we see only the general effect of a number of individually invisible surface features. In fact, however, Mercury's general appearance strikes one as bearing a close resemblance to the way the Moon would look if removed to the same distance, when it would probably display the same lack of sharpness characteristic of all planetary discs.

By analogy, therefore, it seems that Mercury may be very similar to the Moon, at least in its broader characteristics — the extremely rugged nature of the surface and the large grey patches which invite comparison with the lunar 'seas'. If this is so, it may also be possible that the height of the relief is similar, in which case the average height of the mountains would be somewhere between 10,000 and 13,000 feet. However, the most recent observations (by Dollfus at the Pic du Midi Observatory) indicate an upper limit of about 5,000 feet, although the question is still open.

These figures are very different from those suggested by Schröter, who attributed deformations in the southern horn to mountains at least 12 ½ miles high and whose drawings show irregularities at the edge of the planet which, on the scale of the drawing, represent mountains 30 to 60 miles high. These have not been seen since. In all cases, moreover, it is necessary to make proper allowance for diffraction and irradiation, which may easily exaggerate the size of a small white marking, even to the point of making one that is really just near the edge of the disc appear to project from it like a mountain-top.

Active volcanoes, too, have been reported on Mercury. When the planet is passing in front of the Sun as a black disc, there is often a brilliant point of light at its centre, and this point of light has been mistaken for a volcano. The proof that it is illusory is that some observers can see it, while at the same moment others cannot. This

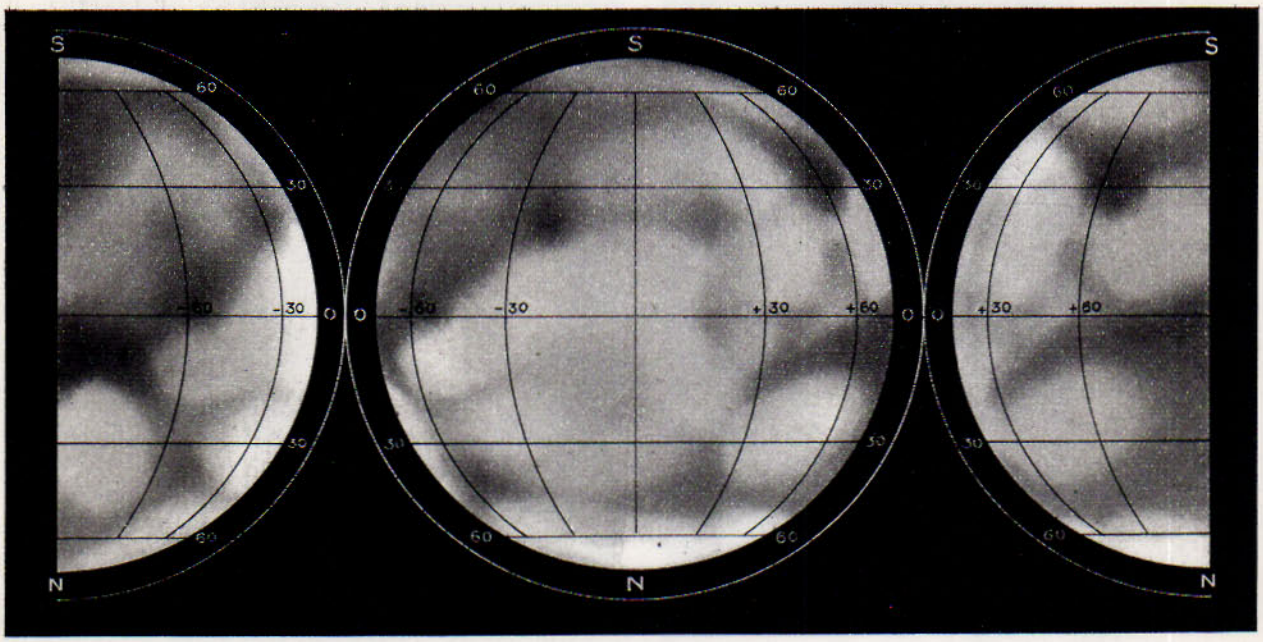


Fig. 357 — CHART OF THE DAY SIDE OF MERCURY. Left: the quarter visible in the morning; right: quarter visible at evening elongations.

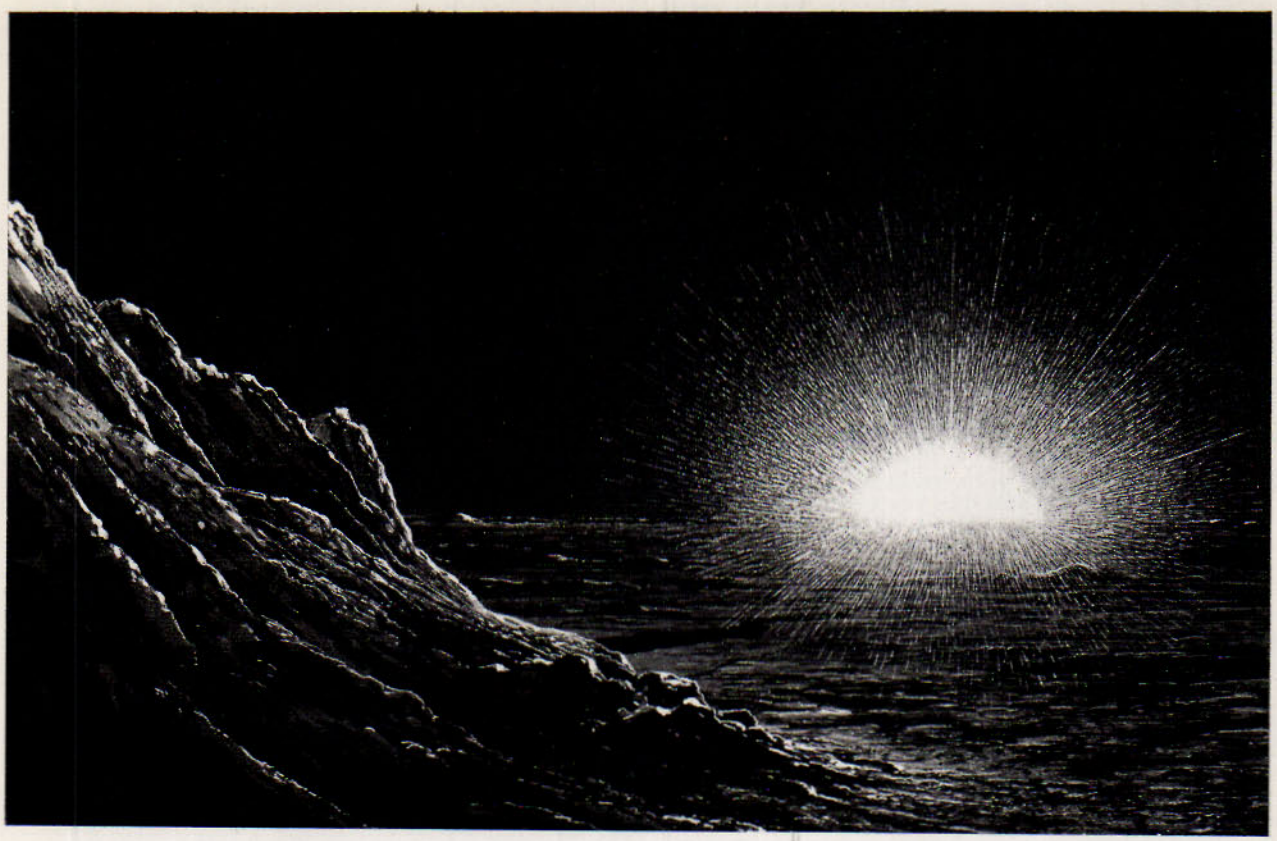


Fig. 358 — SUNRISE AS SEEN BY AN OBSERVER ON MERCURY'S TERMINATOR

shows that it is a purely optical phenomenon caused by diffraction in the instrument used.

Returning to our comparison of Mercury with the Moon, we find the analogy strengthened by the fact that the light reflection or albedo of Mercury is the same as the Moon's. Both bodies as a whole reflect only 7% of the light falling upon them.

And so everything suggests a certain similarity of nature and structure between the two bodies, especially in a complete lack of water. Water, indeed, could not exist in the liquid state on the sunward side of Mercury because of the extremely high temperature obtaining there. It is calculated (and confirmed by radiometric measurements made at Mount Wilson Observatory) that the temperature at the sub-solar point is about 350° C, which is sufficient to melt lead. Water, therefore, cannot exist even as a vapour in the extremely tenuous atmosphere which surrounds the planet.

The atmosphere of Mercury. — When seen in transit in front of the Sun, Mercury frequently appears to be surrounded by a large aureole or halo, sometimes dark but more often very bright. Early astronomers thought that it must be due to an enormous atmosphere.

As in the case of the bright point, this too is an optical effect, this time the result of contrast between the tiny black disc of the planet and the brilliant surface of the Sun behind it.

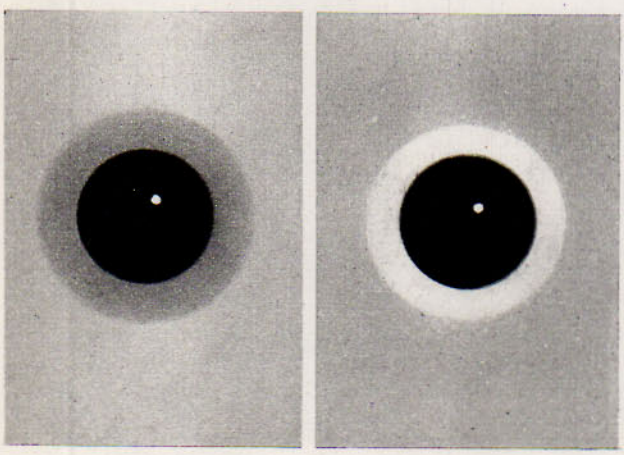


Fig. 359 — AUREOLES OF OPTICAL ORIGIN, AS SEEN BY SOME OBSERVERS DURING A TRANSIT OF MERCURY

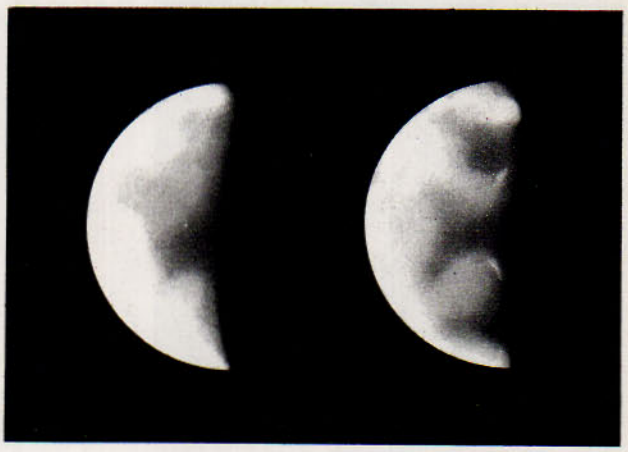


Fig. 360 — TEMPORARY DEFORMATIONS IN MERCURY'S TERMINATOR

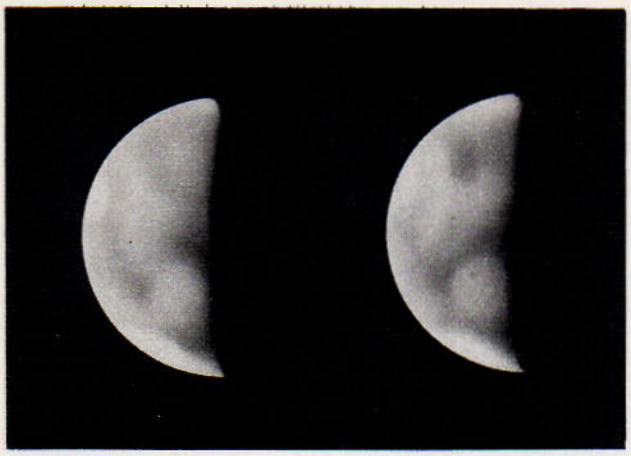


Fig. 361 — VARIATIONS IN THE APPEARANCE OF THE MARKINGS ON MERCURY, SUGGESTING THE PRESENCE OF ATMOSPHERIC VEILS

If Mercury had any atmosphere to speak of, it would clearly betray its presence in various ways. In the first place, it would cause a twilight effect along the terminator, thus prolonging the phase. Secondly, spectroscopic examination gives negative results. Thirdly, photometric investigation shows that Mercury reflects sunlight in the same way as the Moon, which is known to have no appreciable atmosphere.

Thus it seems safe to conclude that Mercury has no atmosphere of any importance. Moreover, theoretical considerations based on the kinetic theory of gases indicate that because of its low mass Mercury could not have retained any atmosphere. Such a conclusion, however, would be at variance with certain observed phenomena which could only be explained on the assumption of some sort of atmosphere, however rarefied it may be.

It appears in fact that all or some of the surface markings disappear at times for a while as if obscured by a whitish veil. This was first noticed by Schiaparelli, and was further investigated by Antoniadi and Rudaux, although in recent years Dollfus, observing at the Pic du Midi, has failed to confirm it. The transient nature of these veils and their movement over the disc could be explained only on the basis of some sort of matter suspended above the ground, i.e. atmospheric disturbances.

Assuming that the obscurations are real, it still remains to discover their origin. They are said to affect certain areas more than others. All the evidence is against their being clouds of water vapour of the type found on Earth, and it seems more likely that they are composed of some sort of extremely fine dust. Despite the tenuity of the atmosphere, such dust could, if once carried into the air, spread over large areas in the form of obscuring clouds. That extremely fine particles can remain suspended in this way is shown by the large quantities of dust expelled into the air by violent eruptions such as that of Krakatoa, which caused red glows at twilight long after.

But with its extremely arid conditions, its extremely tenuous atmosphere, its small size and proximity to the Sun, Mercury is a world essentially different from ours.

VENUS

Elements of the globe. — Of all the planets, Venus most nearly resembles the Earth in size and mass. Its diameter is 7,700 miles; its volume is 0.92 and its mass 0.81 of those of our world. The mean density is 4.9 times that of water, and the polar compression of the globe is so slight that it has not yet been measured.

At first sight, then, Venus and the Earth appear to be almost perfect twins, and it is rather surprising to find that they are very different in nature. Venus proves to be a curious and very hostile world, quite unsuited to any advanced forms of life.

Orbit and movements. — Venus is the second planet, reckoning outwards from the Sun, at a mean distance from it of 67,200,000 miles. This distance is almost constant, since the orbital eccentricity (0.006) is lower than for any other planet in the Solar System. It is also worth noting that, apart from the Moon, Venus is our

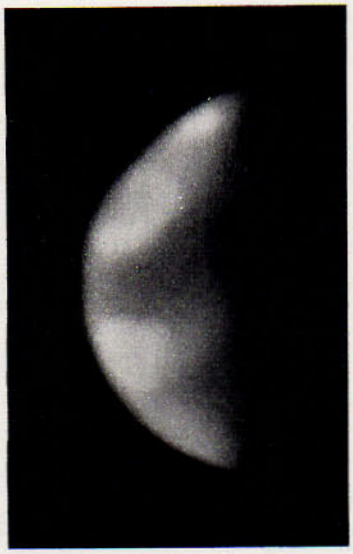


Fig. 362 — ULTRA-VIOLET PHOTOGRAPH OF VENUS, TAKEN BY ROSS AT MOUNT WILSON OBSERVATORY (60-INCH TELESCOPE)

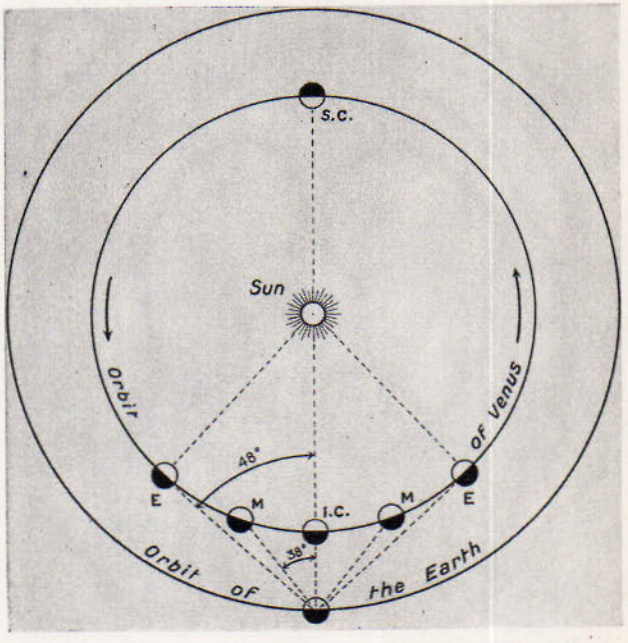


Fig. 363 — RELATIONSHIP BETWEEN THE ORBITS OF VENUS AND THE EARTH

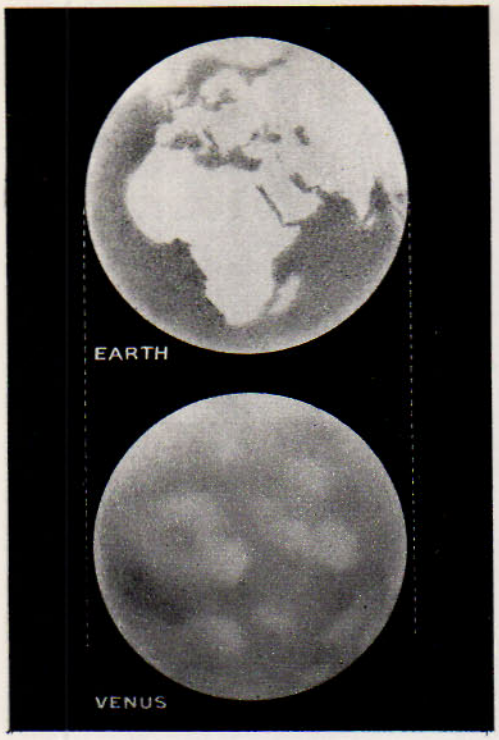


Fig. 364 — THE SIZE OF VENUS COMPARED WITH THE EARTH

Fig. 365 — AT ITS GREATEST ELONGATION, VENUS IS A BRILLIANT OBJECT. Its angular distance from the Sun is greater than 45°.

nearest neighbour; it can approach within 25,000,000 miles, whereas Mars is always more than 34,000,000 miles away.

Since Venus is closer to the Sun than we are, it shows phases similar to those of Mercury. At its greatest elongation (E in Fig. 363) the angular distance from the Sun is some 47°, and Venus is then a glorious object; at its most brilliant it is far superior to any object in the sky except for the Sun and the Moon.

Another fact which may be seen at once from Fig. 363 is that when it is at its closest to us, at inferior conjunction, Venus has its dark side facing us; it lies close to the Sun in the sky, and is difficult to see even with a telescope. As its phase increases, so the apparent diameter shrinks, and when full (superior conjunction) Venus appears as a small disc. The variations in apparent diameter, shown in Fig. 366, range from 65″ to only 9.5″. Obviously, this behaviour presents extra difficulties to the would-be observer of the planet.

The orbital inclination amounts to 3° 23′ 27″, and in consequence transits of Venus across the Sun's face are rare. At the present epoch transits occur in pairs separated by approximately 8 years, with an interval of more than a century between successive pairs. The last transits took place in 1874 and 1882; the next will not occur until 7 June 2004 and 5 June 2012. It is interesting to note that while the 2004 transit will be wholly visible from Britain, only the end of the 2012 transit will be seen from there, since the Sun will rise while the transit is in progress.

The brightness of Venus. — Venus is at its most brilliant during the crescent stage, when its angular distance from the Sun is about 39° east or west — about one month before or after inferior conjunction. Under suitable conditions it may then be seen with the unaided eye in full daylight; when the Sun is well below the horizon, Venus may even cast perceptible shadows. An interesting experiment may be made by placing some objects in front of a *camera obscura* (i. e. a darkened box, without a lens), backed by a photographic plate or film and mounted upon a telescope trained upon Venus as it crosses the sky. After a few minutes it will be found that the emulsion has registered the test object casting its shadow into the back of the box.

The main reason for this great brilliance is that Venus has a very high albedo, amounting to 59%. In fact, what we are seeing is not the true surface of the planet, but merely the upper part of a highly reflective atmosphere.

Telescopic appearance of Venus. — When first observed through a telescope of moderate aperture, Venus appears as an almost perfectly blank disc. Gradually it will be seen that there are lighter and darker patches here and there, but the markings are always so nebulous and indistinct that it is very hard to fix their positions with any certainty. As was realised more than 150 years ago by men such as Schröter and Herschel, the features are purely atmospheric; the actual surface is permanently hidden from us.

The markings are in fact so elusive that doubt has been cast on their very existence. The human eye is always apt to play tricks when straining to catch details at the limit of visibility, and the observer of Venus must be on his guard against optical illusion. To look for the markings when Venus is a brilliant object in the late evening or early morning sky is a waste of time; not only is the disc dazzling, but also the low altitude will make observing conditions hopelessly poor. The only reliable method is to study the planet during full daylight, when Venus is high up.

The markings fall two chief categories: dark shadings and brighter patches. The dark shadings are always ill defined, and seem to change in form and extent over comparatively short periods, so that it is usually difficult to recognise any particular feature from one observation to the next. They behave very much as cloud systems

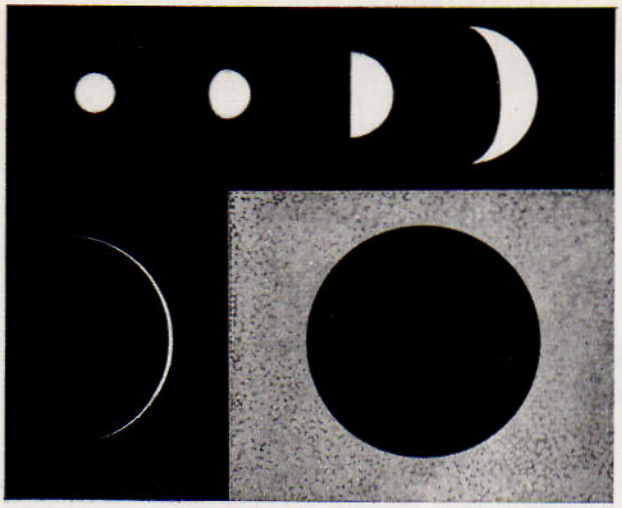


Fig. 366 — VARIATIONS IN VENUS'S APPARENT DIAMETER ACCORDINGS TO ITS PHASE

would be expected to do, and, indeed, it is not misleading to call them simply 'clouds', though they are probably very different from the clouds of our own skies.

Attempts have been made to draw up charts of permanent features upon Venus. As long ago as 1727, F. Bianchini produced a map showing what he believed to be oceans and continents; Percival Lowell, the American astronomer who will always be associated in the popular mind with studies of the Martian canals, drew a detailed chart depicting thin, linear features and circular patches. Up to the time of his death, in 1916, Lowell maintained that the markings on Venus were obvious and constant, but his observations have not been confirmed in recent years, and it is now generally believed that his 'permanent features' were simply optical effects.

During the last thirty years, French astronomers, notably J. Camus, A. Danjon and A. Dollfus, have produced charts showing features which they believe to be always visible under suitable conditions. It has even been suggested that what we are seeing is the true surface of the planet, dimly perceived through the cloudy atmosphere. As will be seen from Fig. 375, there is nothing clear-cut or definite about the markings; but it is fair to say that most observers have grave doubts about any maps of Venus. On the whole, it appears more probable that the shadings are purely atmospheric and impermanent, though the whole question remains open.

Bright patches are seen from time to time, but are no more informative than the dark shadows. All things considered, visual studies tell us depressingly little about Venus as a world.

The rotation of Venus. — One of our main problems concerns the time taken for Venus to turn upon its axis. We know that Mercury has a rotation period equal to its 'year'; the period of Mars has been determined to within a fraction of a second — yet the length of the 'day' on Venus remains unknown.

Many estimates have been made during the last three centuries. In 1666, for instance, G. D. Cassini gave a value of 23h 21m; his son J. J. Cassini amended this to 23 h 20 m, while similar results were obtained by Schröter in 1811 and de Vico in 1841. In 1890, however, Schiaparelli announced that the period must amount to 224 days 16 hours 48 minutes. This is the same length as Venus's sidereal period, and would indicate that Venus, like Mercury, keeps the same face turned permanently towards the Sun.

These and other estimates were obtained by studies of the bright and dark shadings on the disc. Since these features are impermanent and nebulous, the results are bound to be uncertain, and later visual work has done very little to throw light on the problem. Moreover, there is the difficulty that we still have no knowledge as to the

Fig. 367 — VENUS IS SO BRIGHT THAT THE NIGHT SKY CAN BE PHOTOGRAPHED BY ITS LIGHT WITH A ONE-MINUTE EXPOSURE

Fig. 368 — INTENSITY OF VENUS'S LIGHT: a photograph of the shadow cast by a twig, showing diffraction fringes.

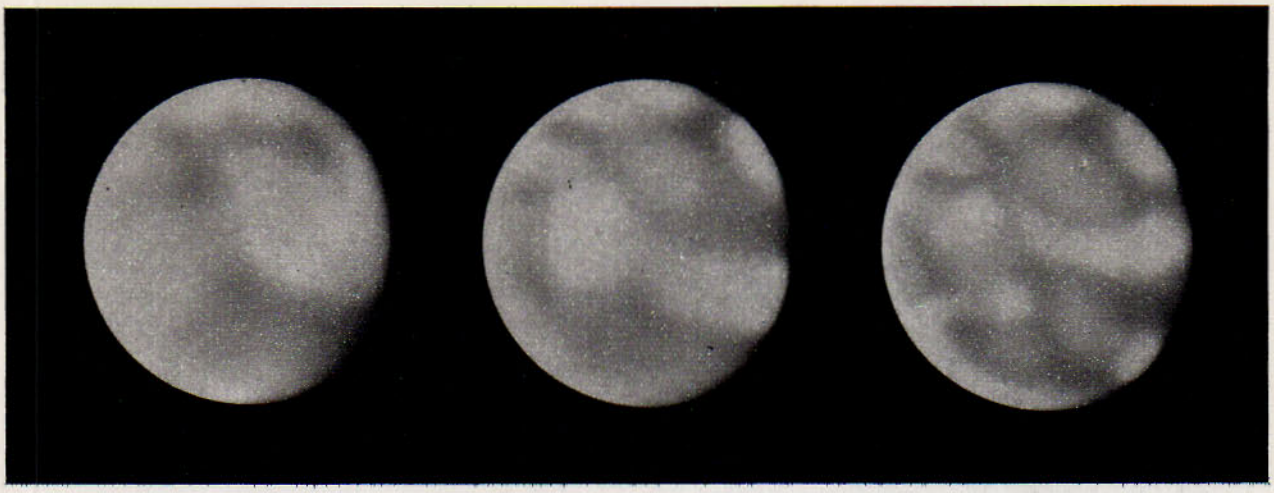


Fig. 369 — THE CHANGING APPEARANCE OF VENUS, DUE TO THE BLOTTING OUT OF DIFFERENT PARTS OF THE SURFACE BY THE THICK CLOUD LAYERS

tilt of Venus's axis. It may well be similar to that of the Earth, a conclusion which is supported by the fact that the horns or cusps of the planet frequently appear brighter than the rest of the disc (Fig. 365); but we have no definite evidence.

In 1900 the Russian astronomer A. Belopolsky attacked the problem by means of the spectroscope. If the axial tilt of Venus is anything like that of the Earth, one limb of the planet will be approaching us, while the other will be receding; this means that the spectrum of the approaching limb will be shifted towards the violet, while the opposite limb will show a red shift. Belopolsky believed that he had detected such effects, and announced a rotation period of 24h 42 m, which he later (1911) amended to 35 h. Unfortunately, spectroscopic work carried out elsewhere led to different conclusions; for instance Lowell considered that his results supported Schiaparelli's very long rotation period.

It was hoped that photography might settle the question once and for all, but again the results were disappointing. Photographs of Venus taken in ordinary light, or in infra-red, show practically no details of the disc, which is an extra proof — if proof were needed — that what we see is simply the upper part of an atmospheric layer. In 1927, however, F. E. Ross at Mount Wilson obtained an excellent series of photographs in ultra-violet light, one of which is reproduced as Fig. 362. Definite features could be made out, and from them Ross deduced a rotation period of about 30 days, though with great uncertainty.

In 1956 J. D. Kraus, of Ohio, announced that he had detected radio waves coming from Venus. He suggested that they might be due to electric phenomena in the planet's atmosphere, perhaps not unlike our thunderstorms, and from fluctuations in their strength he derived a rotation period of 22h 17m. Kraus gave a maximum uncertainty of only 10 minutes either way, but it cannot yet be said that his result is at all conclusive, particularly since no other research team has managed to detect any radio emanations from Venus.

So far the problem remains unsolved. During the past few years several detailed investigations have been carried out, with results that are hopelessly contradictory; in addition to Kraus's value of 22h 17 m, G. P. Kuiper has suggested 'a few weeks', while Dollfus and other French astronomers continue to support the 225-day period. R. S. Richardson, at Mount Wilson, has made another attempt to determine the period spectroscopically, and considers that the true value must lie somewhere between 3 ½ and 46 days.

The atmosphere of Venus. — In 1761 Venus passed in transit across the face of the Sun. This transit was observed by the Russian astronomer M. V. Lomonosov, who detected a slight blurring or haziness of the planet's disc, and stated that 'Venus is surrounded by a considerable atmosphere equal to, if not greater than, that which envelops our earthly sphere'. The existence of an atmosphere was fully confirmed later in the eighteenth century by Schröter, who correctly assumed that the shady markings were of a cloudy nature.

Since Venus is so like the Earth in size and mass, it is logical to suppos ethat its atmosphere, too, must be similar. Yet this is not the case. Spectroscopic analysis by Nicholson, St John and others has shown that there is no detectable water vapour or free oxygen, and that the atmosphere is extremely rich in carbon dioxide gas.

Carbon dioxide exists in the Earth's air-mantle, and is also to be found in that of Mars, but the quantities are relatively small. The fact that it is so abundant on Venus leads to some important conclusions. For instance, carbon dioxide has a strong 'greenhouse' effect; it tends to blanket in the Sun's heat, and so the surface temperature of Venus is likely to be high. Moreover, an atmosphere so rich in carbon dioxide seems to preclude the existence of any form of advanced life.

The nature of the 'clouds' which hide the surface of Venus so effectively is still a matter for dispute. Ross, following his 1927 photographs, suggested that the upper atmosphere might be impregnated with yellowish dust blown from the surface by violent winds, while ten years later R. Wildt put forward the theory that the clouds are due to formaldehyde, a combination of carbon, hydrogen and oxygen (CH_2O) formed under the influence of ultra-violet light emitted by the Sun. Another theory, that of H. Suess, suggests the cloud layer is composed of salts such as sodium chloride and magnesium chloride, produced by the drying up of former oceans on the planet.

In 1954 two American astronomers, F. L. Whipple and D. H. Menzel, carried out a long study of the problem, and reached some rather startling conclusions. They pointed out that a series of measurements of the polarisation of the light of Venus made in 1929 by Bernard Lyot did not fit in at all with the dust-cloud idea; indeed, among the substances with which he experimented, Lyot found that only water droplets agreed reasonably

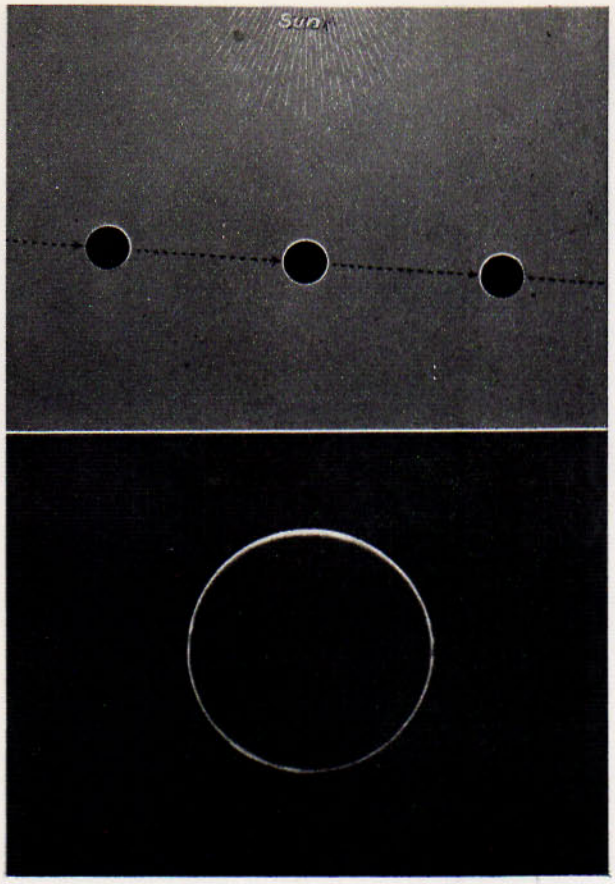


Fig. 370 — VENUS'S ATMOSPHERE CAN SOMETIMES BE SEEN AS A BRIGHT RING

with the variation of polarisation with scattering angle on Venus. Whipple and Menzel stressed that the temperature of the planet's upper atmosphere is low — the most recent estimates, by J. Strong and W. M. Sinton, give about −40°C — and that under such conditions there would be little hope of detecting water vapour. The final conclusion was that the clouds are made up of ordinary water droplets.

It is still too early to make any definite statements, and it is clear that our knowledge of the character of Venus's atmosphere is deplorably incomplete.

The Ashen Light. — We have referred previously to the fact that when the Moon appears as a crescent it is often seen that the 'night' side is shining dimly. Strangely enough, a similar effect has also been reported with regard to Venus, and this so-called 'Ashen Light' is not nearly so easy to explain. Venus, as will be shown later, has no satellite, and the Earth would be quite unable to cause any perceptible glow upon the darkened hemisphere.

The Ashen Light was first recorded in 1643 by Riccioli, a Jesuit professor at Bologna. Since then it has been seen by almost every regular observer of Venus, but again we must beware of optical illusion; contrast effects may play their part, and some authorities believe the Light to be a purely optical phenomenon.

If the Light is real, and is not due simply to contrast, the most plausible explanation is that it is due to electrical disturbances in the planet's atmosphere; it may in fact be similar to the aurorae which form so striking a feature of the far northern and southern skies of Earth. N. Kozirev, at the Crimean Astrophysical Observatory, claims that he has obtained spectra of the Light, and that there are indications of ionised nitrogen — also conspicuous in the spectra of aurorae. Once again, however, confirmation is lacking.

The 'Satellite of Venus'. — The Earth has one large satellite, the Moon. Since Venus is almost a match for our world in size, it, too, might be expected to have a satellite, and only two hundred years ago the existence of a moon circling round Venus was regarded as established. Astronomers such as G. D. Cassini, J. Short and Mayer had reported it from time to time, while in May 1761 Montaigne, of Limoges, made a series of observations of it which seemed conclusive.

Yet this satellite of Venus has not been glimpsed since 1764, and there seems to be no doubt that it is nonexistent. The bodies described by Short, Montaigne and the rest were either faint stars or (more probably) optical 'ghosts' of the type which may be expected with a very brilliant object such as Venus.

It is not impossible that there may be a minute satellite a few hundred yards in diameter, but there is no proof. So far as we know at present, Venus, like Mercury, is unaccompanied.

The surface of Venus. — Direct observation can tell us nothing about the real surface of Venus; at every turn we are defeated by the all-concealing atmosphere. However, it is at least possible to make some reasonable guesses.

The great quantity of carbon dioxide in the atmosphere leads to a high surface temperature, and the lack of free oxygen is some indication that plant life is not abundant. This point of view is fairly recent; less than thirty years ago, before much was known about the atmospheric

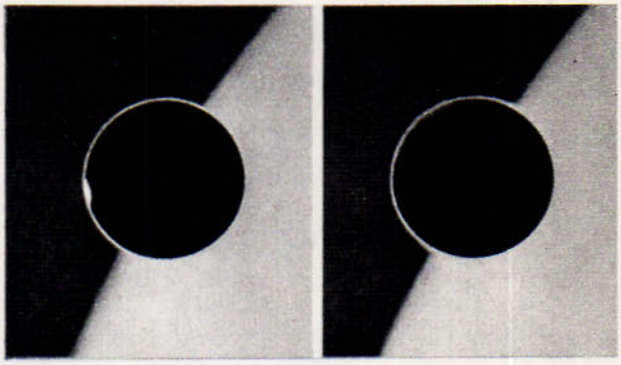


Fig. 371 — THE ATMOSPHERE OF VENUS, OBSERVED DURING TRANSITS OF THE PLANET. Left: after Russell, 1874; right: after Langley, 1882.

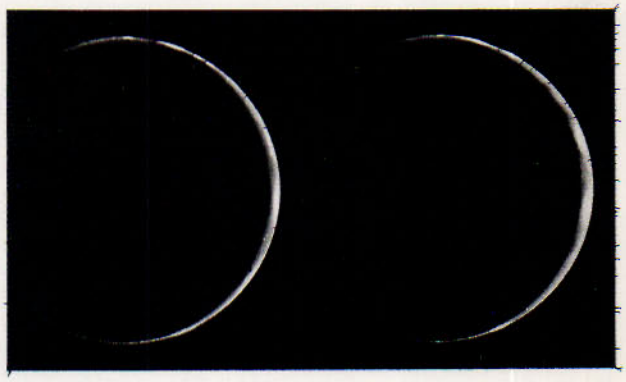


Fig. 372 — PROLONGATIONS OF THE CRESCENT DUE TO THE ILLUMINATION OF THE ATMOSPHERE. *(L. Rudaux)*

Fig. 373 — To an observer at high altitude, a sea of clouds masking the surface of the Earth gives an impression of the way in which Venus's surface is generally hidden by the turmoil of the atmosphere.

composition, it was generally thought that Venus must be a world in a state similar to that of the Earth perhaps 250 million years ago, when the coal deposits were being laid down. According to this rather attractive idea, Venus was swampy and moist, with luxuriant vegetation of a primitive type, and probably insects and amphibians too.

As soon as the abundance of carbon dioxide was realised, the 'coal forest' Venus had to be abandoned, to be replaced by a Venus that was a dusty desert, without a drop of moisture anywhere. This is still the theory favoured by many astronomers, and is indeed reasonable enough. Whipple and Menzel, on the other hand, think that the entire surface is covered with water. The truth of the matter is that we still simply do not know, and whether it will be possible to find out within the next few years remains to be seen.

Life on Venus? — Though our knowledge of Venus is so fragmentary, all the evidence shows that men could not exist there; indeed, animal life in any form appears to be out of the question. Some Russian astronomers, notably G. A. Tikhoff, consider that there are wide areas of orange or red vegetation, but there is no proof that plants exist at all. It is far more likely that Venus is a world where life has not developed; though there is always a chance that if the Whipple-Menzel 'water theory' is correct, the oceans of Venus may contain primitive creatures of the sort which flourished in our own seas more than five hundred million years ago.

MARS

Elements of the globe. — Compared with the Earth, Mars is only a medium-sized planet; it has a diameter of 4,200 miles, which is only just over half of the Earth's. This means that its surface area is 3.6 times and its volume 7 times smaller than those of the Earth.

There is also a considerable difference between their masses, on account of the lower density of the materials composing the Martian globe. Their density is estimated at 3.8, in contrast to the Earth's 5.52. From this it can be calculated that the force of gravity is much smaller at the surface of Mars than here: an object weighing 10 lb. on Earth would only weigh 3.7 lb. on Mars. Similarly, a freely falling object would fall only about 6 feet in the first second, as opposed to 16 on Earth.

The planet is very slightly ellipsoidal in shape and the flattening, which has not yet been measured with complete accuracy, is estimated to be 1/200. Its rotation period, on the other hand, is known with great precision: 24h 37m 22.56s. This means that the Martian days and nights are only a few minutes longer than on Earth. There is yet another point of similarity: the inclination of Mars's axis is between 24° and 25°, so that the seasonal variations closely resemble those of the Earth, whose axial inclination is almost the same (23° 27′). Consequently, at similar latitudes on Mars the same long days will be followed by the same short nights in summer as on Earth, and vice versa in winter. The Martian seasons, however, are almost twice as long as the terrestrial ones and vary more in length. This is due both to the planet's revolution period, which is twice as long as ours, and to the higher eccentricity of its orbit.

This is important not only because of the physical conditions it produces on the surface of the planet, but also because of the great variations it causes in the visibility of Mars from the Earth.

Fig. 374 — TWO DRAWINGS OF VENUS, showing the possible configuration of its surface through gaps in the cloud-layers. Left: 10 May 1929; right: 28 September 1935. *(L. Rudaux)*

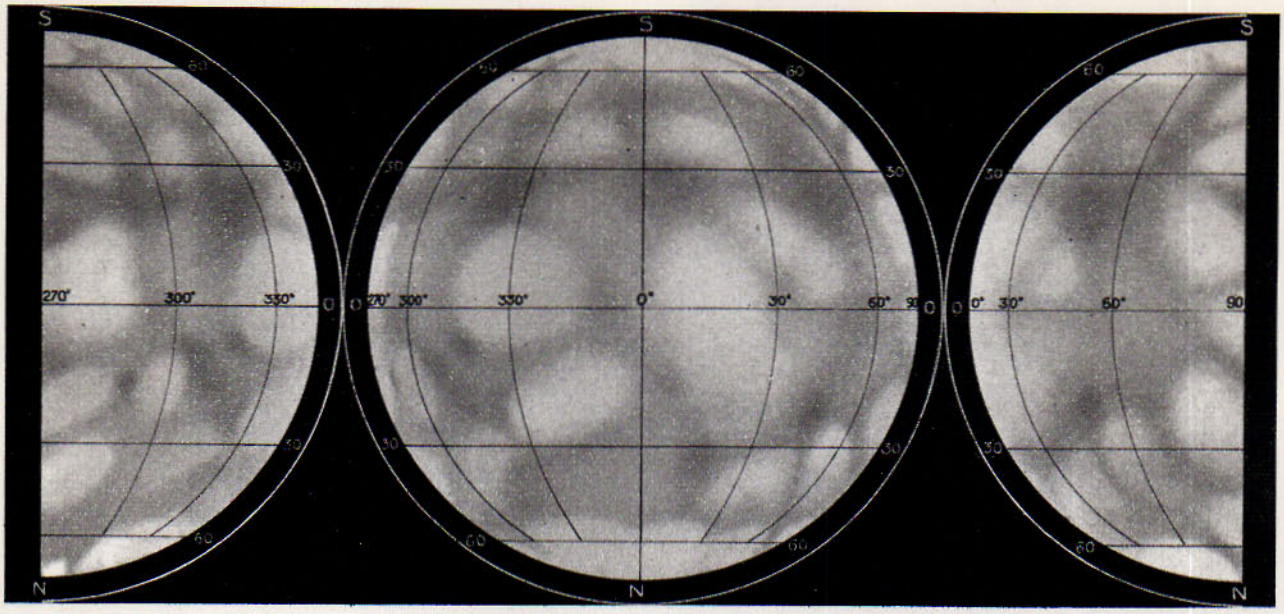


Fig. 375 — CHART OF THE ILLUMINATED HEMISPHERE OF VENUS. Left: quarter visible in the morning; right: quarter visible in the evening.

Orbit and motion of Mars. — Because of the marked eccentricity of the Martian orbit (0.093) the planet's distance from the Sun undergoes considerable variations: from 128,750,000 miles at perihelion, it increases to 155,000,000 miles at aphelion. On account of its shape, therefore, the orbit of Mars (Fig. 378) is far from being concentric with that of the Earth. Where these two celestial highways approach most nearly, they are only 35,000,000 miles apart, whilst at their maximum separation the distance between them is 62,000,000 miles. Since the Earth and Mars move at different speeds, the point where two planets are in opposition does not always occur in the same place relative to the position where the two orbits are nearest together; the opposition distance of Mars therefore varies continuously. This is very important in relation to the observation of very small details on Mars, since the apparent size of the disc depends on its distance. We shall now explore this matter more fully.

With a mean orbital velocity of 15 m.p.s., Mars makes one complete revolution in 687d 23h. Since Mars moves more slowly than the Earth, the two bodies only come into line again after 780 days. This period of 780 days equals two complete revolutions of the Earth, plus an extra 50 days, corresponding to one-seventh of its orbit.

Let us start with the two planets in opposition and with Mars at perihelion, so that it is at its closest to the Earth. Because of the difference in their speed, the next opposition will take place a little further round, one-seventh further round, in fact. Owing to the eccentricity of the two orbital ellipses however, this will mean that the two planets are now further apart than at the first opposition. This trend will continue until the two planets come to opposition at their maximum opposition distance. From then on the order is reversed and finally they return to their starting point. The complete cycle takes approximately 15 years, during which there are 6 oppositions between the 2 perihelic oppositions.

This explanation is given only in general terms and is merely designed to show the mechanism whereby the distance between the Earth and Mars varies at each opposition. In actual fact the cycle is not repeated exactly each time since the revolution period of Mars is not a multiple of the Earth's, and so oppositions do not recur at exactly the same place. Perihelic oppositions, for example, do not always occur when Mars is at its absolute minimum distance; they come a little to one side or other of this point, but never, or hardly ever, miss it by exactly the same amount. The same holds true for the other types of opposition — aphelic and intermediate.

Furthermore, it is also necessary to take into account the constant variations in distance during any particular apparition (the time between two successive conjunctions with the Sun). During this time, Mars is constantly changing its position relative to the Sun and Earth, owing to the difference between the orbital velocities of the two planets. Fig. 378 shows that the greatest possible separation is when Mars, in conjunction with the Sun, is at aphelion. Thus the distance between Mars and the Earth, allowing for every conceivable combination of relative positions, varies between 35,000,000 and 249,375,000 miles.

Leaving aside the differences in the apparent size of the disc which these variations introduce, the angle at which Mars is tilted towards or away from the Earth is also constantly varying. This is due to the inclination of its axis of rotation. As we have already discussed the mechanism responsible for the seasons, we shall now merely show how it applies in the case of Mars.

The equinoxes and solstices are so arranged in relation to the major axis of the planet's orbit that for each Martian hemisphere the four seasons are of very unequal length (Fig. 380). Because opposition with the Earth may occur at any season on Mars, it quite frequently happens that one of the poles is tilted steeply towards us whilst the other is hidden on the far side of the disc. Sometimes, too, both poles appear exactly on the limb.

How we observe Mars. — When Mars is at its minimum distance from the Earth, the planet's apparent diameter reaches its maximum value of 25.1″; when at its furthest, Mars is only 3.6″ in diameter. This inequality, however, is not always present between opposition and conjunction; as we have seen, the eccentricity of their orbits is such that the distance between the two planets will vary even when their position relative to the Sun is the same. For the same reason, Mars's apparent diameter also varies from one opposition to the next all through the fifteen-year cycle. Fig. 381 shows diagramatically

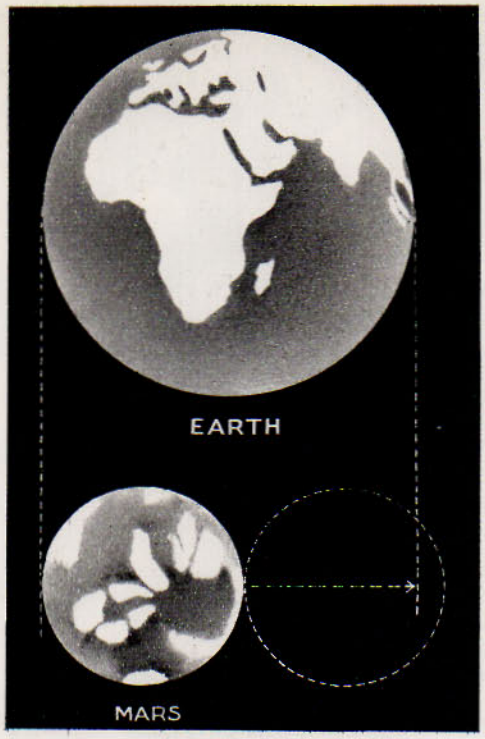


Fig. 376 — THE SIZE OF MARS COMPARED WITH THE EARTH

how this occurred for the cycle lasting from the perihelic opposition in 1939 to the last one, in 1956.

The axis of Mars is inclined in such a way that the south pole is exposed at perihelic oppositions and the north pole at aphelic ones. Because the apparent diameter of the Martian disc is greater at perihelion than aphelion, it is the southern hemisphere of the planet which is presented under the most favourable conditions, and which can therefore be examined in the greatest detail.

This will explain the differences between photographs of Mars taken at different times.

Lastly is should be noted that at certain times the angle between the Earth, Sun and Mars is such that we can see a little way round the dark side of the planet, and the disc shows a perceptible phase-effect. This is very slight, however, since the width of the unilluminated segment is never more than one-seventh of the diameter of the disc at quadrature.

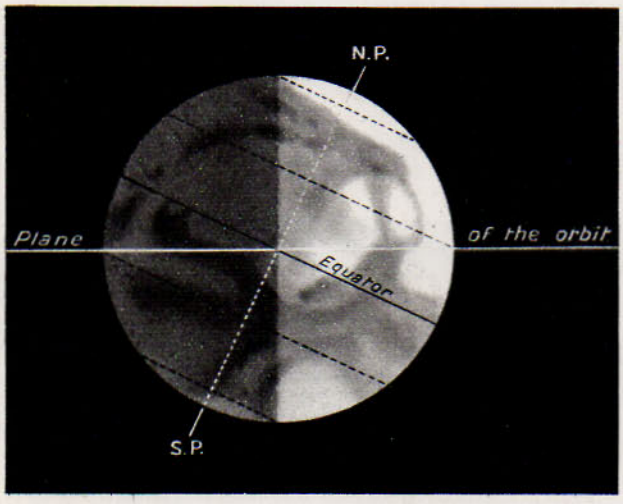


Fig. 377 — INCLINATION OF MARS'S AXIS

The surface appearance of Mars. — Of all the planets in the Solar System, Mars is the one about which our knowledge is most certain. It is true that Mars never approaches the Earth as closely as Venus, but on the other hand Mars has the advantage of being much the easier to observe, because when at its closest, at opposition, it presents a fully-illuminated disc, whilst Venus is either a thin crescent or else completely invisible.

With the naked eye, Mars is a splendid sight; it is clearly orange in colour and has been compared with a glowing coal, or a jewel. This quality is not quite so pronounced in the telescope, which shows a disc covered with markings which range from an ochre or yellowish tint to a yellowish pink, and a more ruddy colour in some places. Standing out clearly in contrast to this are the dark markings, which are grey or slightly green. This pattern of markings is permanent, i.e. its essential features can always be made out, although their visibility depends on the seeing conditions and the telescope used.

In the main, however, these markings are plain enough for them to have been seen right at the start of real telescopic observation, despite the poor quality of the images given by the instruments of those days. The fact that the markings could always be definitely made out does not mean that they have always been drawn in the same way on the various maps that have appeared. Although the outlines of the coarser features are quite easy to see, the same thing cannot be said of the smaller details of the picture. In this connection, it should be remembered what was said at the beginning of this chapter about the delicate nature of telescopic images, and the way in which every observer gives a slightly different interpretation of the same subject.

Having made this proviso, let us now examine the surface of Mars in greater detail; for this task we require fairly powerful optical aid, especially if our interest lies in the finer features of the surface structure. Once a very rudimentary science, observation of Mars has become more and more refined through the years, with the growing power of telescopes and the constant improvement

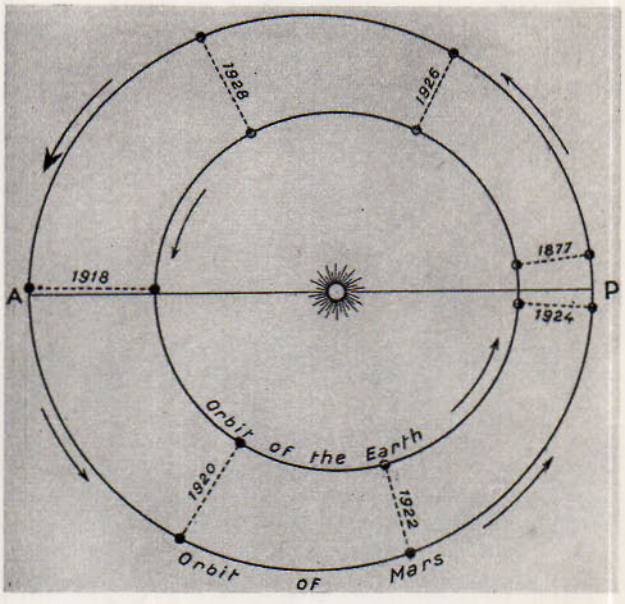


Fig. 378 — RELATIONSHIP BETWEEN THE ORBITS OF MARS AND THE EARTH

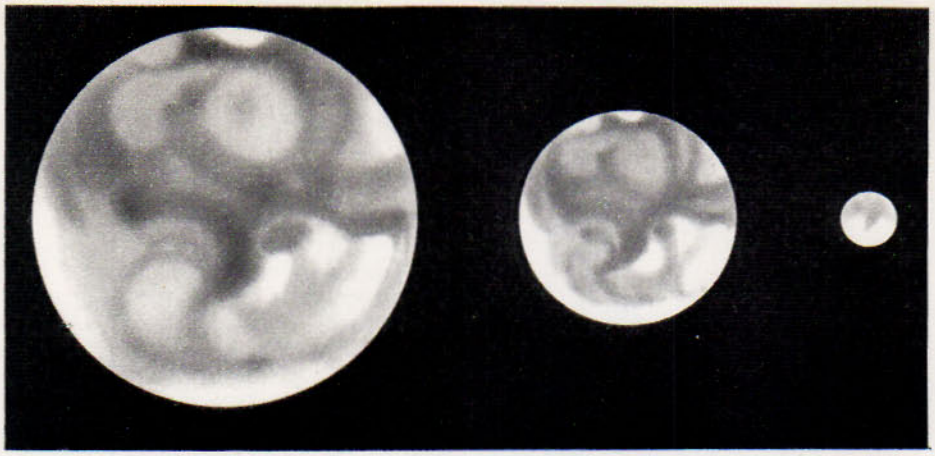


Fig. 379 — VARIATIONS IN THE APPARENT DIAMETER OF MARS. Left: perihelic oppositions; centre: aphelic oppositions; right: conjunctions.

in their optical quality. The result of this is that most of the ideas which were once accepted have now given way to new ones, often directly opposed. There has, indeed, been a constant shift away from the original theories, which considered Mars simply a smaller version of the Earth.

Another very characteristic feature consists in the two white patches which occupy the polar regions of the planet, and which are bright enough to be seen in a fairly modest instrument. These very obvious polar caps undergo regular variations in size, which are closely correlated with the progress of the seasons on the planet. These will be discussed more fully later.

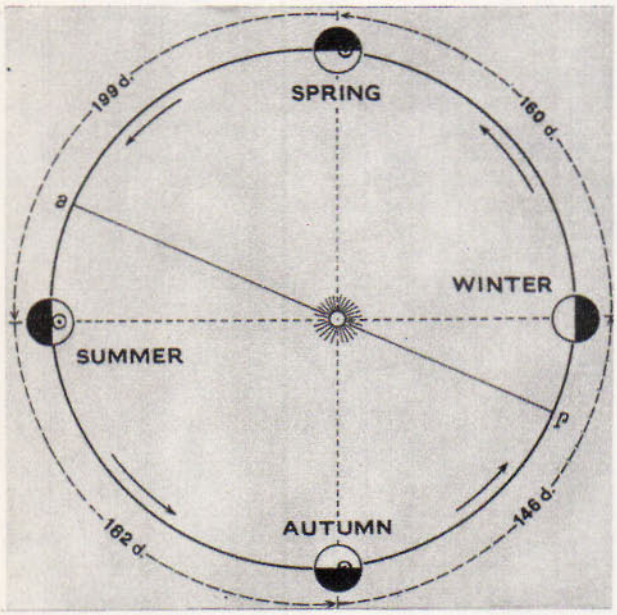


Fig. 380 — LENGTH OF THE MARTIAN SEASONS

Through the telescope one's first impression of Mars is usually of a planet covered with oceans and continents like our own: the bright yellow or orange areas may be thought of as the continental regions, whilst the grey-green patches corerespond to the oceans. This is what the early observers thought, and from it has arisen a whole system of nomenclature as misleading as that applied to the Moon: continent, land, island, ocean, sea, strait, etc., are all used in reference to different areas of the planet. This nomenclature is due mainly to Proctor, whose famous chart, compiled in 1867, summarised all the information then available.

When this chart is compared with a modern one, it is at once obvious what tremendous strides have been taken in the observation of Mars during the last century or so. The same general pattern in still there, of course, but it cannot be denied that the early observers interpreted complex features in a very over-simplified manner, and vague markings were frequently drawn with a hard outline which has no objective existence. This mode of drawing was no doubt fostered by the preconception that Mars was covered with land and sea, the hard line presumably being thought of as a coast. Observations made with powerful modern instruments of the highest quality do not substantiate this interpretation.

What emerges from modern work is not this simple pattern of oceans and continents, with their sharply defined shore-lines, but a remarkable complexity of structure, containing many differences in tonal intensity, particularly in the shades of colour, which vary considerably. This

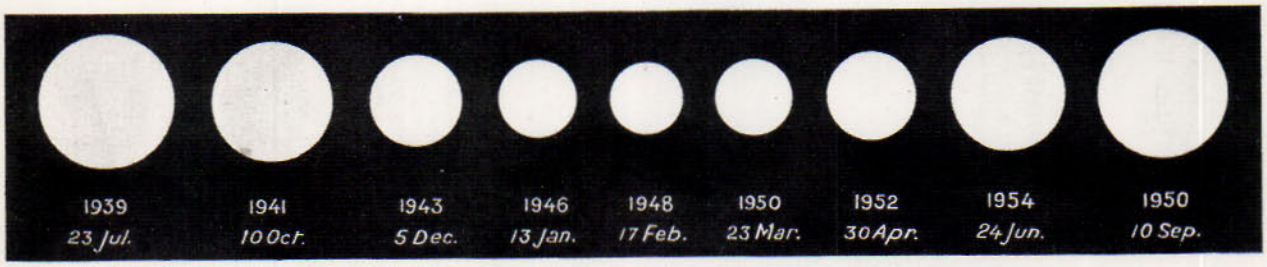


Fig. 381 — VARIATIONS IN THE APPARENT DIAMETER OF MARS AT EACH OPPOSITION DURING THE CYCLE LAST COMPLETED: from the perihelic opposition of 1939 to that of 1956.

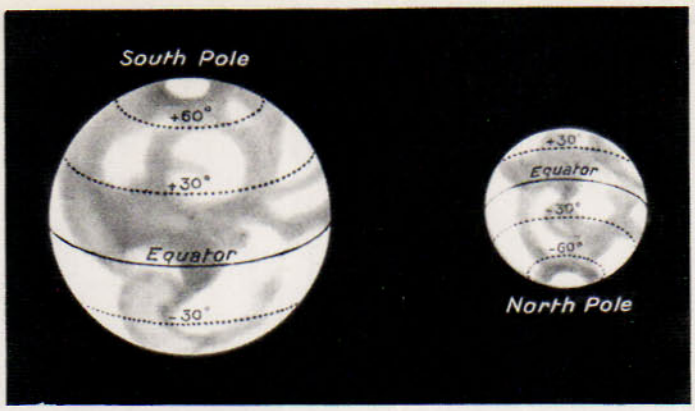


Fig. 382 — MARS'S APPARENT DIAMETER AND THE TILT OF THE GLOBE.
Left: at perihelic opposition; right: at aphelic opposition.

last point will be discussed later, since there are a number of interesting questions involved. Considering simply the intensity of the markings, we find that they show great differences in darkness, with some very dark patches among the half-tones and brighter spots. This explains to a certain extent why the markings themselves and their exact outlines are not always represented in the same way; their appearance depends partly on the conditions under which they are observed. In spite of this, it seems well established that real variations do take place, some occasionally being of such magnitude as to cause a profound alternation in the appearance of a large area.

The most written-about, spoken-about, and thought-about features on Mars are undoubtedly the curious rectilinear dark markings which have been named 'canals'. Before discussing this controversial topic in detail, however, let us first survey the main facts about Mars which have been accumulated up to the present time.

The atmosphere, the polar caps and the climate. — We know that Mars has an atmosphere since certain observations can only be explained on this basis — e.g. hazes, which often spread over large areas and hide the surface from view, deposits of frost, etc. This is confirmed by ultra-violet photography.

Investigations along various lines have not always resulted in perfect agreement, and the most recent work suggests that conditions on Mars are somewhat different from what had long been accepted as the true picture. In fact, the question under discussion is not whether Mars actually has an atmosphere, but how much and of what sort.

Fig. 383 — AT OPPOSITION MARS SHINES BRILLIANTLY AMONG THE STARS
(Photo, L. Rudaux)

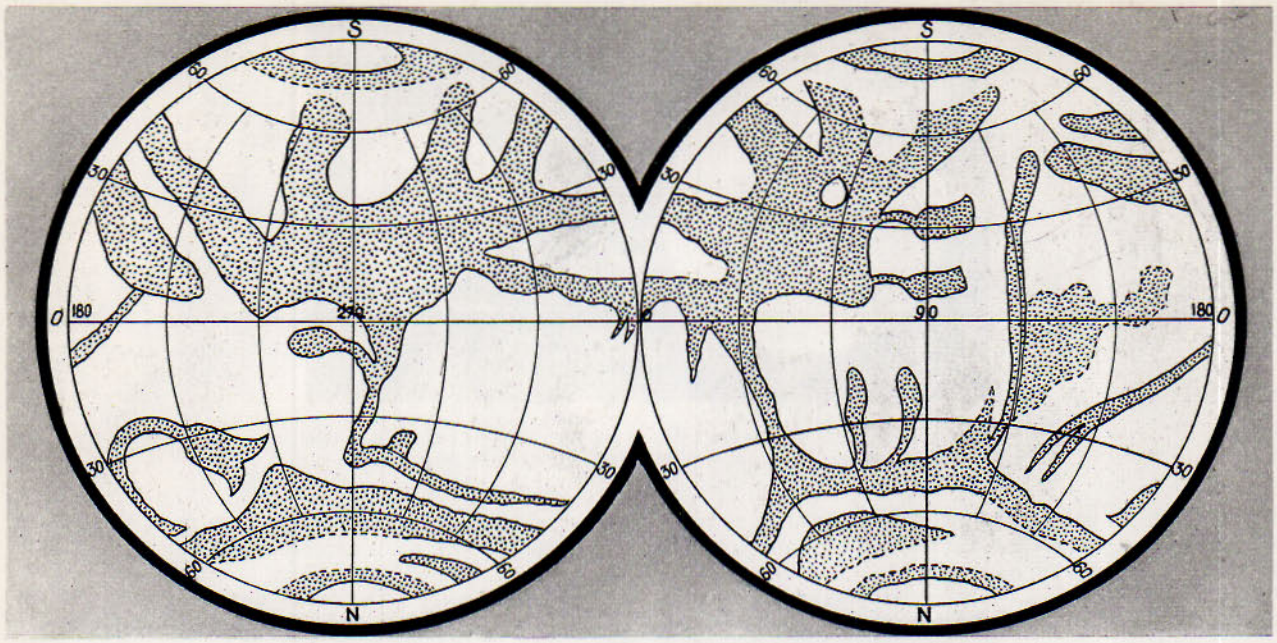


Fig. 384 — PROCTOR'S CHART OF MARS (1867)

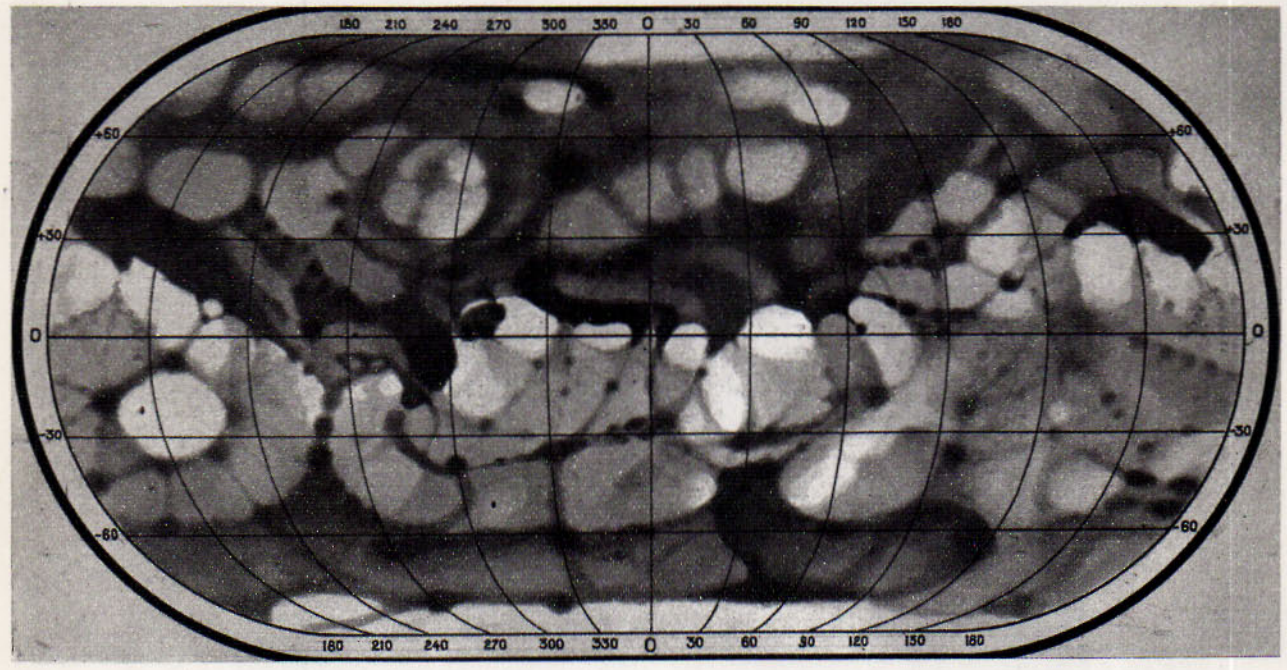


Fig. 385 — ANTONIADI'S CHART OF MARS, GIVING THE RESULTS OF MORE MODERN WORK ON THE PLANET

The principal evidence in favour of an atmosphere is the presence of the polar caps. They grow during the Martian winter and then diminish considerably or vanish entirely in spring and summer. It is safe to say that they must be deposits of ice, snow, or more probably frost, and the fact that they regularly change in size according to the season argues in favour of an atmospheric circulation, complete with condensation and melting or sublimation, according to the temperature.

These seasonal variations are visible in quite small instruments, and they take place on a relatively large scale. This is shown by the fact that the polar caps may extend as much as 25° or 30° from the poles in winter, only to contract with the onset of spring, so that by the end of summer, when the whitish veils of haze begin to appear, they have become extremely small or have disappeared entirely. During autumn and early winter, the polar cap re-forms beneath its whitish covering of haze. The amplitude of these seasonal variations is not the same in each hemisphere because of the unequal length of the seasons:

Southern spring or northern autumn	lasts			146 days;
— summer	—	winter	—	160 days;
— autumn	—	spring	—	199 days;
— winter	—	summer	—	182 days.

The effects of this inequality are exaggerated by the fact that for the southern hemisphere, aphelion occurs in winter and perihelion in summer. Consequently the

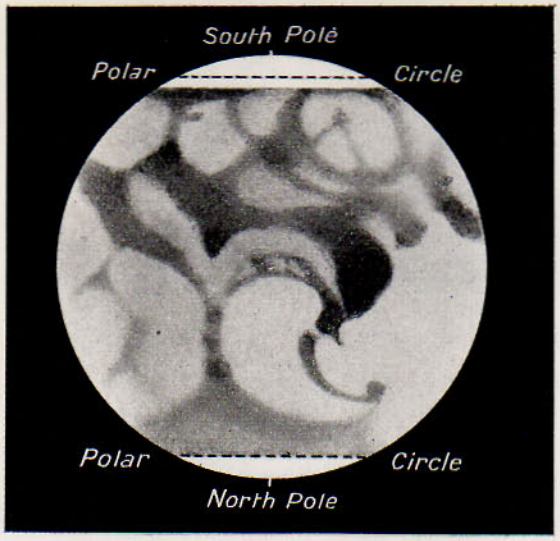


Fig. 386 — THE COMPARATIVE EXTENT OF THE NORTH AND SOUTH MARTIAN POLAR CAPS

amplitude of the variations in the size of the south polar cap is greater than for the northern.

In addition to the purely astronomical conditions involved, it seems probable that local physical conditions play some part in the waxing and waning of the polar caps. In certain places the white mantle of frost or snow retreats much more rapidly than in others, whilst elsewhere the whiteness lingers when surrounding areas have been uncovered. This suggests the presence of high ground and valleys, and the only such area we have been able to detect with certainty on Mars can be seen near the south pole as a block of white detached from the rapidly dwindling cap. It is a plateau of sorts, but cannot be much more than about 2,500 feet high.

Such considerations as these suggest that early explanations of the nature of the polar cap were not entirely correct. It was assumed that they were composed of great sheets of ice, in the same way as those on Earth. This idea, however, has had to be abandoned, since it can be shown that such great quantities of ice would not be supplied with sufficient heat from the Sun to allow them to melt away completely during the Martian winter. Indeed, being so much further away from the Sun, Mars receives considerably less heat than the Earth. All we know about the Martian atmosphere suggests that abundant precipitation is impossible, which means that these extremely thick ice-caps would not be able to re-form.

All this nevertheless gives evidence for the existence of an appreciable Martian atmosphere. A second sign is the formation of a dark band around the polar caps when they are melting in spring. This band is several hundred miles across and it follows the retreating ice-cap. Although the existence of such a band was for some time a matter of debate, it is now well established as an objective phenomenon. It undoubtedly represents a belt of country which has been dampened by the retreating ice, and the fact that this moisture remains in the liquid state for even a short time indicates that the minimum atmospheric pressure on Mars must be at least a few centimetres of mercury.

Further proof of the existence of a Martian atmosphere is furnished by the frequently observed hazes which spread over large areas obliterating the well-known surface features. Among observations of this sort must be mentioned the masses of whitish cloud, which bear a vague resemblance to terrestrial cloud systems in the general way in which they move. Furthermore, clouds are occasionally seen projecting from the terminator; to be visible in this way they must lie at a height of about 10 or 20 miles. Ultra-violet photography reveals the existence of widespread hazes, which are not normally visible in white light, but become so if very intensely bright.

To these hazes, which behave rather like our cirrus (although it is impossible to say how alike they really are) must be added disturbances of another sort. The orange-yellow colour of Mars was once attributed to atmospheric phenomena, and although this is not entirely true, it does appear that clouds of this colour form fairly frequently. This fact was first recognized by observers at the end of the 19th century, and has been confirmed by Antoniadi, Fournier, de Vaucouleurs and others. These coloured veils can cover vast areas of the planet, obscuring the surface features, changing the colour of the markings on the ground, and sometimes giving the polar caps themselves a reddish tinge. This sort of thing occured as recently as the 1956 opposition, when very little of the well-known markings could be seen.

The height of these clouds does not seem to be very

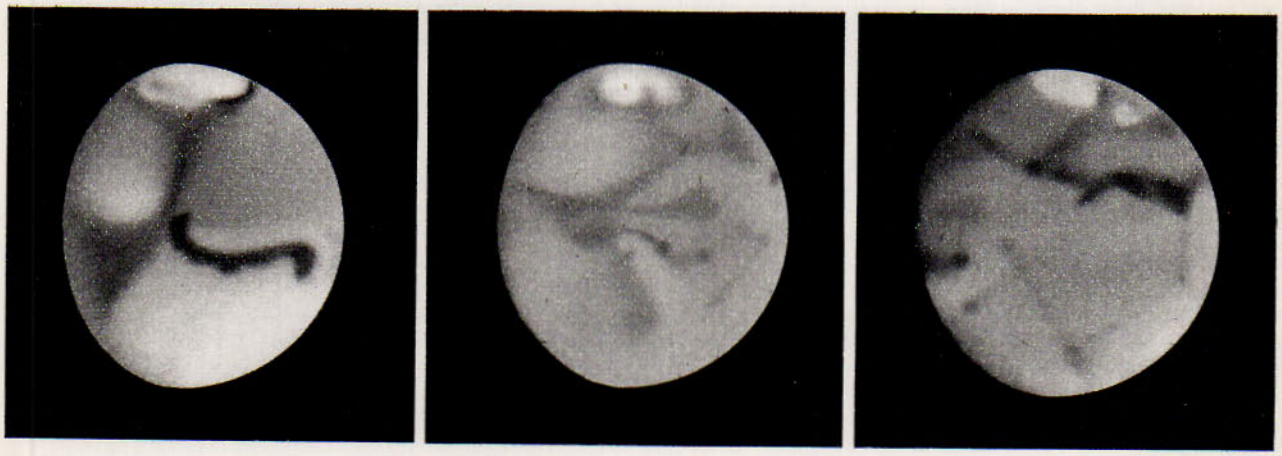


Fig. 387 — CHARACTERISTIC BEHAVIOUR OF MARS'S SOUTH POLAR CAP AND ATMOSPHERE. Left: 17 July 1941, 1h 28m U. T. The polar cap is bordered by a dark belt, which has completely disappeared in the views from a month later. Centre: 11 August 1941, 0h 43m U. T. The surface features, seen through light atmospheric hazes, are very faint. Right: 11 August 1941, 4h 27m U. T. Four hours later the hazes cross the terminator, revealing their height by the slight bulge they produce. *(Drawings by G. de Vaucouleurs, Houga Observatory, Gers)*

Fig. 388 — PHOTOGRAPHS OF MARS TAKEN IN MONOCHROMATIC LIGHT. Left and centre: in ultra-violet light only the planet's atmosphere and the clouds can be seen. Right: in infra-red the actual surface appears plainly. *(Photographs by R. Trumpler, Lick Observatory, 1924)*

great — about 3 miles. According to Antoniadi, they are most prominent when the planet is at perihelion, i.e. when the radiation received from the Sun is at its maximum. His explanation of these climatic vicissitudes seems plausible: he suggests that they are clouds of dust raised in the Martian deserts by the comparatively strong winds at that time of year. The reasons for believing that deserts exist on Mars will be given later, since they are connected with our ideas about the size and importance of the atmosphere.

Following certain observations, of the polar caps amongst others, it was once thought that the atmosphere of Mars must be very similar to our own. Early spectroscopic work, too, indicated the presence of large quantities of water vapour on Mars, although even then its atmosphere was recognised as being much more transparent and much less cloud-laden than our own. Except for the last point, modern work has shown the situation to be very different. By various astrophysical methods, such as photographic and visual photometry, and by investigation of the way in which Mars polarises the light it reflects back into space, it has been possible to make a tentative estimate of the atmospheric pressure at the planet's surface, and values of a few centimetres of mercury have been found. This means that on Mars the pressure is almost one-twentieth of the normal atmospheric pressure at the Earth's surface; alternatively, it is equal to the pressure obtaining in the stratosphere at a height of about 15 miles.

At all events, this pressure is sufficient for water to exist in the fluid state on the surface of Mars, which would substantiate the theory that the dark bands observed around the polar caps are melted ice.

On the other hand spectroscopic work carried out with very powerful instruments by Dunham and Adams at Mount Wilson observatory has failed to detect the presence of either water vapour or oxygen in the atmosphere of Mars. More accurately, it enables us to say that if there is any oxygen in the Martian atmosphere, it must be in quantities well under 1/100 of those in which it is found on Earth, while the amount of water vapour must certainly be much less than 1/500 its terrestrial value. The presence of carbon dioxide in the Martian atmosphere, on the other hand, is definitely established, thanks to Kuiper's observations of 1948.

Although these values are, of necessity, approximate, they do serve to underline the differences between the atmospheres of the Earth and Mars, and confirm the advanced state of dehydration reached by Mars, whose meagre reserves of water gather round the poles every winter in the form of frost or snow.

These results agree perfectly with the radiometric measurements of the temperature of different parts of the planet's surface made at Flagstaff and Mount Wilson.

From this work, it can be deduced that the mean temperature of Mars, as is natural considering its great distance from the Sun, is appreciably lower than that of the Earth: about $-20°$ to $-30°$C, compared with $+10°$ to $+15°$C on our own planet.

In the Martian tropics, however, the highest temperatures are well above zero, reaching $+10°$ to $+20°$ in the bright regions and $+20°$ to $+30°$ in the dark patches, at mid-day. These values are for perihelion; at aphelion, when the planet is much further from the Sun, the temperature hardly ever exceeds 0°C.

Towards evening, the temperature begins to decline rapidly, and at sunset is in the neighbourhood of 0°C. There is a further fall in the night, probably down to $-40°$C, for at sunrise the recorded temperatures are still only about $-20°$ to $-30°$. These enormous diurnal extremes of temperature give proof of the tenuity of the Martian atmosphere, and its low water vapour content. (In this connection we may well recall our earlier account of the April Moon on page 175.)

The above information relates to the tropical and equatorial regions. It is obvious that in the temperate zones the temperature must be proportionately lower, and in winter, during the long polar nights, the prevailing temperature will be nearer $-80°$ to $-100°$C, although rising to as much as 0°C in summer.

Thus we can say that although Mars is the planet which differs least from the Earth, it will severely test any human beings landing there, and they will need all of and more than the equipment now used by high-altitude flyers to sustain life.

The nature of the Martian surface. — In the days when it was thought that the markings on Mars were oceans and continents, certain observers began to suspect that changes were taking place. Amongst these was Father Secchi of the Vatican Observatory, whose work was confirmed by Camille Flammarion from 1876 onwards. To account for these changes, it was suggested that the dark markings corresponded to shallow seas with ill-defined shores, resembling the large tidal swamps that are found in the tropics on Earth. This would explain many apparent changes in the extension of certain markings, assuming some change, however small, in the level of the oceans.

This view of Mars was held for many years. Within the 'land areas', it was thought, logically enough, that in the absence of any structural details there must be watercourses, lakes, valleys, hills covered with luxuriant vegetation (since the characteristic reddish tint of these regions was considered by many people to be due to some sort of plant life). Moreover, a certain monotony in

Plate III. THE PLANET MARS

the landscape was generally admitted, because of the failure to detect high ground of any importance on Mars. Indeed, as far as we can tell, Mars is completely devoid of high ground (with the exception of the low plateau near the south pole), since mountains of only a few thousand feet in height would be detectable under suitable conditions from Earth, appearing as deformations in the terminator.

All these considerations apply to the reddish or ochre regions of the planet, which were once thought to be continents. The dark patches, which were then taken for oceans, were assumed to be diminishing in size owing to the drying-up of the water. Nowadays we know that what open water Mars ever had has long since disappeared. How then do we interpret the dark patches today?

It was stated earlier that the dark patches were of a rather greenish colour; some of this is no doubt due to contrast between them and the reddish colour of the rest of the disc. Minute examination of these markings, however, shows that they are made up of an assembly of smaller structural details; they almost give the impression of a mosaic, which, when seen from a distance, takes on the appearance of a blurred mixture of colours and tones resulting from the fusion of smaller structural elements.

Many of these dark markings frequently change their colour. According to Antoniadi, some remain green or blue; others turn from green to brown during spring and summer, whilst some change from grey to carmine or lilac brown. All this is incompatible with the nature of oceans.

The most remarkable fact, however, is that these changes are related to the seasons. One hypothesis immediately springs to mind: that the ground is covered with living organisms which change their appearance under the influence of changing temperature and humidity. By analogy with our own world, we say that these organisms are vegetable in nature, but we must take care not to push the analogy too far, and say, as some have done, that we are observing woods and meadows and the like; this conjures up a vision of grass, bushes and trees such as we know on Earth, which is quite foreign to Mars. It shows that we must be careful, even in the choice of words.

This caution is justified by the study of photographs of Mars taken in infra-red light. A green terrestrial landscape photographed in this way looks as if winter had suddenly arrived, for the trees and fields appear to be covered in snow. Furthermore, since infra-red is not much affected by the atmosphere, details which the latter usually prevents our seeing are clearly revealed. Masses of vegetation, which cannot normally be seen against the ground because of their distance, show up unmistakably as patches of white in the photograph.

When photographed in this way, the image of Mars

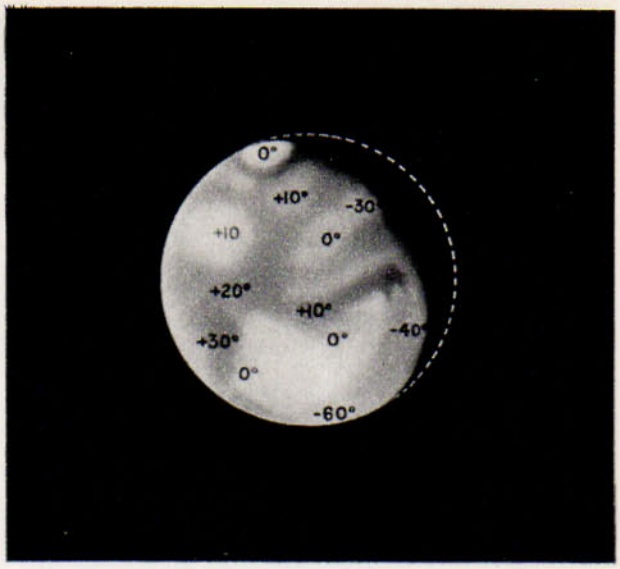


Fig. 389 — TEMPERATURE DISTRIBUTION ON MARS AT THE START OF THE SOUTHERN SUMMER, 1926. *(Measurements made by W. W. Coblentz, Lowell Observatory)*

shows no comparable effects. The dark surface markings look exactly the same in infra-red as in other wavelengths of light: if anything they are slightly more pronounced. If there is no special and unknown factor at work here we can safely say that from the facts before us, the regions on Mars which are supposed to be covered with vegetation do not behave in infra-red light in the same way as terrestrial vegetation. They cannot therefore contain the green chlorophyll which gives green plants their characteristic appearance in infra-red, and this implies a totally different biology for the Martian 'plants'.

If, in spite of this, the Martian vegetation, if such it is, sometimes behaves like terrestrial vegetation, we may well call to mind the example of our swamp lands on Earth. There the varying amount of water supports an appropriate type of vegetation-covering; furthermore, over much larger areas than the stretches of open water, the same conditions exist because of the general dampness of the ground. Similarly, this is how, in the midst of deserts, there can exist oases of green vegetation, drawing their water from underground sources.

If the nature of the main Martian surface markings can be explained in this way — which does not preclude the possibility that areas of open water may exist in places — the problem of explaining most of the variations is solved, at least in theory. The humid regions are, in fact, liable to all sorts of variations depending on the seasonal

Fig. 390 — VARIATIONS IN THE APPEARANCE OF A SMALL AREA OF MARS: THE LACUS SOLIS, as seen by Schiaparelli in 1877, and by Antoniadi in 1924 and 1926.

Fig. 391 — A LANDSCAPE PHOTOGRAPHED IN INFRA-RED LIGHT

climatic changes. This applies equally to the visibility of the canals, which will be discussed later. The canals are very probably some sort of special surface feature resulting from the action of tectonic forces causing an alignment of a number of depressions, which, because they collected moisture, became relatively well adapted to support life.

In one way and another, these curious features are very interesting, because they give what is the first recognisable evidence for the existence of life on another world. Although the precise nature of the dark markings on Mars is still under discussion, the nature of the ochre areas which cover about three-quarters of the planet's surface can now be decided with considerable accuracy.

Photometric observations show that the proportion of sunlight diffused by these ochre areas, 15 to 20%, is in every way comparable with that of our sandy deserts, so that, like them, they probably owe their characteristic ruddy colour to the oxide of iron known as 'rust'. Moreover the polarimetric measurements made by Bernard Lyot at Meudon Observatory suggest that the surface of the planet is of a dusty nature and this agrees with the theory that the reddish hazes are in fact sandstorms. Finally, the American physicist W. W. Coblentz has detected absorption bands attributable to the silicates in the planet's infra-red spectrum, which would reinforce the analogy between — if not the identity or nature of — the Martian and terrestrial deserts.

The only difference, but an important one, resides in those dark lines which have been christened with the unfortunate name of 'canals'.

The canals of Mars. — The curious canals of Mars were not really noticed until the opposition of 1877, when Schiaparelli detected them from the Observatory of Milan. He pursued his observation intensively and systematically during the next two oppositions and continued until giving up all visual work when his sight began to fail. His observations show an astonishing system of lines of varying intensity, linking one dark spot to another, crossing the 'continents', and covering the surface of the planet with a sort of geometrical network.

It may be noted, however, that some of the more prominent canals appear on earlier drawings: e.g. by

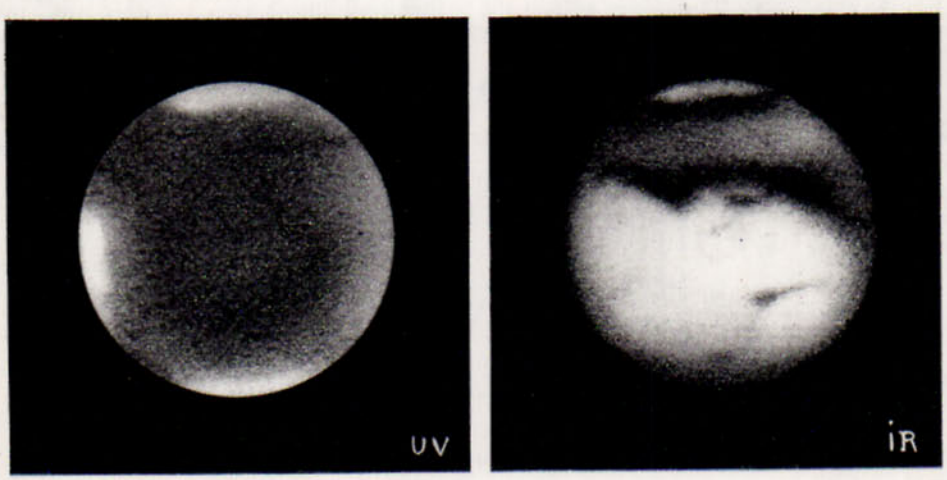


Fig. 392 — MARS PHOTOGRAPHED IN ULTRA-VIOLET (LEFT) AND INFRA-RED LIGHT, 29 July 1939. *(After W. H. Wright, Lick Observatory).*

MARE CHRONIUN
Tiphys Fretum
Chersonesus
Electris
Eridania
Ausonia
Zea Locus
HELLAS
MARE AUSTRALE
ARGYRE I
Dia
Aonius Sinus
Phaethontis
Icaria
MARE ERYTHRÆUM
Bosphorus
THAUMASIA
Dædalia
Memnonia
Mesogæa
Amazonis
Arcadia
Tempe
Noachis
Pandoræ Fretum
Iapygia
Deucalionis Regio
Margaritifer Sinus
Eos
Auroræ Sinus
Juventæ Fons
Lux
Cimmerium
Hesperia
Tyrrhenum
Adriacum
Zephyria
Æolis
Antæus
Cerberus
Lestrygon
Sirenum
Ausonia Borealis
Libya
Syrtis Major
Ætheria
Cebrenia
Utopia
Phelgra
Elysium
Isidis Regio
Neith Regio
Umbra
Copais Palus
Dioscuria
Cydonia
Edom
Aram
Oxia Palus
Oxia
Aeria
Arabia
Phison
Euphrates
Hiddekel
Eden
Gehon
Indus
Chryse
Xanthe
Nilokeras
Tractus
Albus
Ismenius Lacus
Deuterchilus
Protonilus
Mare Acidalium
Nix Tanaïca
Ceraunius
Mare Boreum
Lunæ Lacus
Ascræus Lacus
Nix Olympica
Castorius Lacus
Propontis
Mare Sirenum
Adonis
Electris
Phœnicis Lacus
Memnonia
Mesogæa
Arcus
Tritonius Lacus
Nectar Lacus
Solis Lacus
Coprates
Candor
Nodus Gordius
Nodus Altyaicus
Mercurii Insula
Coloe Palus
Hadriacus
Thoth
Hermæa Syrtis
Pambotis Lacus
Triviæ Chaos
Cerberus
Lestrygon
Sinus Eumenides
Sinus Sabæus
Sinus Meridiani
Sinus Asopus
Sinus Boreum
Sinus Fontis
Sinus Juventæ
Sinus Auroræ
Sinus Margaritifer
Sinus Serpentis
Sinus Cimmerium
Sinus Eumenides
Sinus Sirenum
Sinus Phœnicis
Sinus Lux
Sinus Tritonius
Sinus Nectar
Sinus Solis
Sinus Coprates
Sinus Candor
Sinus Albus
Sinus Ascræus
Sinus Nix Olympica
Sinus Castorius
Sinus Propontis
Sinus Adonis
Sinus Electris
Sinus Phaethontis
Sinus Icaria
Sinus Dædalia
Sinus Memnonia
Sinus Mesogæa
Sinus Arcus
Sinus Amazonis
Sinus Arcadia
Sinus Tempe
Sinus Nix Tanaïca
Sinus Ceraunius
Sinus Dioscuria
Sinus Cydonia
Sinus Utopia
Sinus Cebrenia
Sinus Ætheria
Sinus Elysium
Sinus Phelgra
Sinus Isidis Regio
Sinus Neith Regio
Sinus Umbra
Sinus Copais
Sinus Palus
Sinus Oxia
Sinus Aram
Sinus Edom
Sinus Eden
Sinus Aeria
Sinus Arabia
Sinus Phison
Sinus Euphrates
Sinus Hiddekel
Sinus Gehon
Sinus Indus
Sinus Chryse
Sinus Xanthe
Sinus Nilokeras
Sinus Tractus
Sinus Albus
Sinus Ismenius
Sinus Deuterchilus
Sinus Protonilus
Sinus Mare Acidalium
Sinus Mare Boreum
Sinus Mare Australe
Sinus Mare Chroniun
Sinus Mare Erythræum
Sinus Mare Hadriacum
Sinus Mare Tyrrhenum
Sinus Mare Cimmerium
Sinus Mare Sirenum
Sinus Mare Hellespontus
Sinus Mare Serpentis
Sinus Mare Eumenides
Sinus Mare Sabæus
Sinus Mare Meridiani
Sinus Mare Asopus

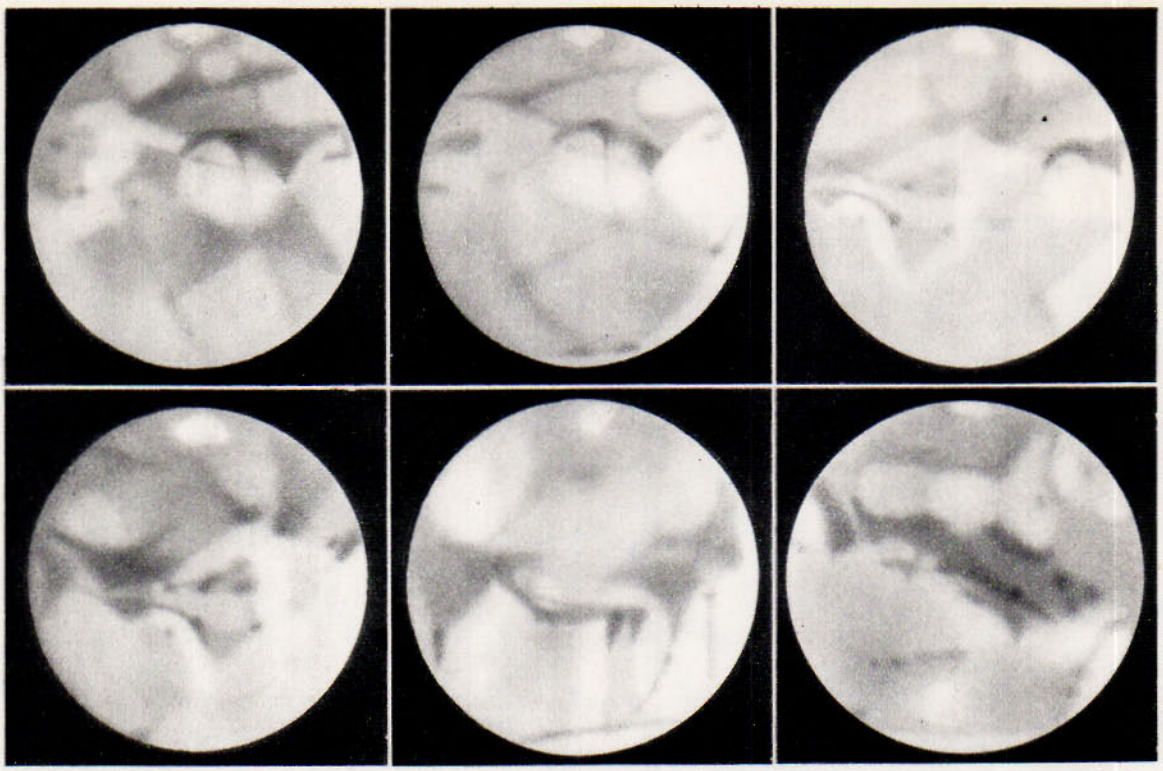


Fig. 394 — SUCCESSIVE VIEWS OF MARS, SHOWING ONE COMPLETE ROTATION OF THE GLOBE
1. 11 September 1941 at 0h 43m; U.T. Long. 131°. 4. 18 September 1941 at 0h 44m; U.T. Long. 70°.
2. 12 — — 1h 56m; — — 141°. 5. 23 — — 22h 34m; — — 344°.
3. 15 — — 0h 36m; — — 94°. 6. 6 October — 22h 45m; — — 128°.
Drawings by G. de Vaucouleurs, Houga Observatory. 8-in. refractor, magnification × 260. Note the streaks, termed 'canals', visible on drawings 1, 2, 3 and 5. Compare with the photographs on the opposite page taken from the Pic du Midi at the same time.

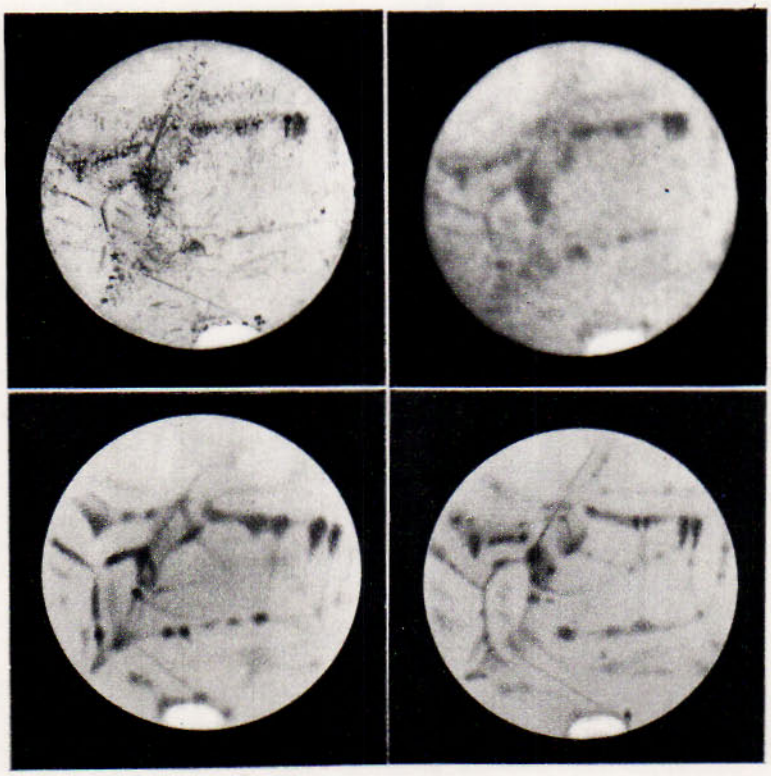


Fig. 395. A LABORATORY EXPERIMENT WITH MARS. A test drawing (top left) covered with fine details roughly resembling the appearance of Mars was drawn (lower two pictures) by two experienced observers under the same sort of conditions as those obtaining during real observations. The test object was also photographed with the same resolving power and under the same conditions. Comparison of the four views reveals the subjective element which enters into all attempts to observe Mars, especially its canals.

20 September, 2h; 70° 18 September, 2h 20m; 93° 18 September, 2h 35m; 97°

19 October, 23h 20m; 126° 15 October, 22h 35m; 151° 12 October, 22h 25m; 178°

16 October, 2h 20m; 206° 10 October, oh 25m; 231° 8 October, oh 40m; 252°

8 October, 3h 58m; 300° 1 October, 3h 48m; 359° 27 September, 3h 34m; 31°

Fig. 396 — A PHOTOGRAPHIC TOUR OF MARS
Composité pictures resulting from the superimposition of short-exposure images photographed by B. Lyot, H. Camichel and M. Gentili, with the 15-inch refractor of the Pic du Midi Observatory, at the opposition of 1941.

Schröter in 1798, and more particularly by Dawes in 1864. But it was the work of Schiaparelli which first called attention to them, on account both of their interesting distribution and of the way in which the great Italian astronomer drew them. At the same time Schiaparelli drew up a general chart of the planet, which showed the topography with a precision and richness of detail which had never before been attained, and in the place of the old system of names he substituted the now accepted Latin nomenclature, taken from terrestrial geography, mythology and classical antiquity generally.

After Schiaparelli, the canals were seen by a number of observers, one of the most assiduous being Percival Lowell, the famous American astronomer, who in 1894 founded an observatory at Flagstaff, Arizona, specially devoted to the study of Mars. More and more canals were seen, and the observers at Lowell Observatory began to publish drawings covered with an incredibly tangled network of lines. Even greater interest attended the announcement that some of the canals displayed the phenomenon of gemination, or doubling: that is to say, in place of a single line, two appeared side by side, like a railway track.

Since all such observations were strenuously disputed by other astronomers, such announcements provoked some controversy. To give some idea of the atmosphere at this time, let us review the main facts of the case.

The canals, which cut the Martian surface in all directions, have most frequently been described as very delicate features, and so regular that on many drawings they have been shown as thin lines. Others, however, have been drawn as hard or indefinite bands of considerable width, so that they are more immediately obvious.

Whatever their nature, these formations seem to occupy quite definite positions, that is to say, are connected with certain quite definite surface structures, which they appear to link together, for the canals are frequently found along the line where two patches of differing shadow intensity come together. But it may be noted that various observers report that these curious lines continue across the great dark surface shadings. Moreover, where two canals intersect in one of the bright regions, they enlarge to form the small, round spots which have been given the name 'oases', and oases sometimes occur along the length of a canal like beads on a string. All this led Lowell to his famous theory that the geometrical network of canals was the work of intelligent 'Martians', irrigating the parched soil of their rapidly drying planet.

Not all of these features, it may be noted, are equally prominent, and they appear to undergo certain variations. It is rather difficult to be absolutely positive about the last point, because when examining such minute details so much depends on the quality of the telescopic image, and the precision and accuracy of the observation.

The artificial nature attributed to the Martian canals has rightly excited men's curiosity for a long time. But it must be realised that this view was deduced from certain observations, and that observations are liable to error, as are observers. In fact, when the most powerful modern optical resources are brought to bear on the problem, under the best conditions, the highly artificial regularity disappears, and it can be seen that the 'canals' are composed of small mottlings and dark patches on the surface, arranged roughly in lines. When observed with inadequate optical power, these individual patches remain just on or below the threshold of vision, and show up as blurred, continuous streaks, in the same way as a chain or string of beads will take on the appearance of a single straight line when seen from a distance.

Observations of this type made by Antoniadi, using the 33-inch refractor of Meudon Observatory, would seem to be conclusive. But his view, natural though it seems, has not been universally accepted.

Fournier, one of a number of gifted observers, has pointed out that certain canals which are usually so faint as to be of doubtful nature, sometimes develop into quite unmistakable broad, dark bands; the opposite has been known to occur when prominent canals disappear completely.

These observations have been confirmed by the present members of the Lowell school. Slipher in particular has *photographed* a number of the more prominent canals. and oases, single and double, even though they arefar from being the hair-like lines which Lowell's drawings would seem to indicate.

Finally, admirable photographic and visual results have been obtained since 1941 at the Pic du Midi Observatory high up in the French Pyrenees, first by Bernard Lyot and, after his untimely death, by Audouin Dollfus. Their work clearly reveals the great structural complexity of the canals and the surface generally. It also emphasises how critical the atmospheric conditions on Earth must be, since the slightest trembling of the image blurs all the minute details, and the rectilinear canal-like appearance stands out sharply.

Early interpretations of the nature of these uniquely Martian surface features seem to have been based

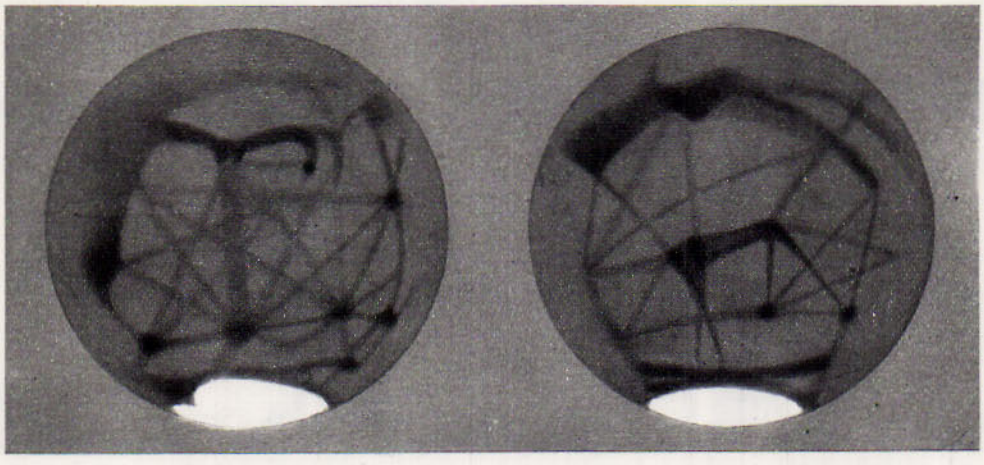


Fig. 397 — Drawing of Mars showing the geometrical aspect of the 'canals'. After Douglass, Lowell Observatory, 1898.

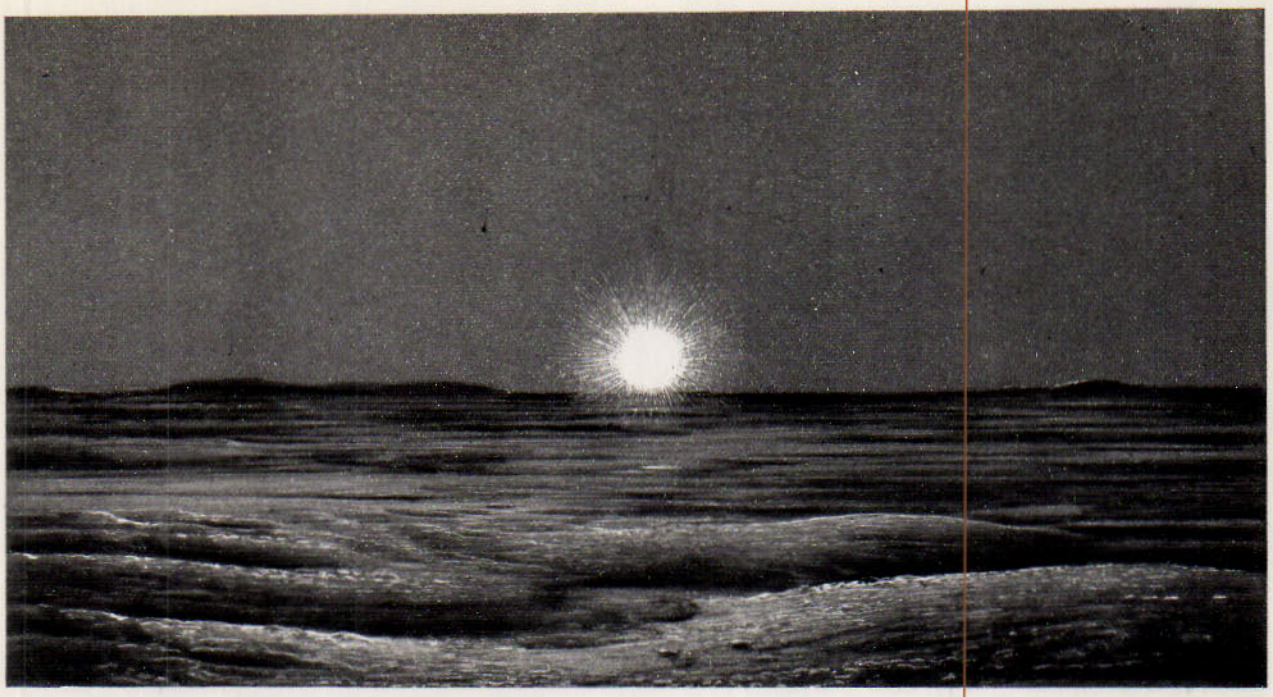


Fig. 398 — SUNRISE ON MARS

on a false interpretation of the evidence, and on over-simplified observations which showed smooth features where the surface is actually very diversified, so that these ideas have had to be considerably modified during recent years. However, there does seem to be a modicum of truth in them, and they cannot be rejected out of hand. Although the canals are neither artificial waterways nor bands of irrigated land on either side of such waterways — as Lowell thought — we still have to explain why these particular surface features are arranged in linear patterns. This aspect of the canal controversy is still an open question, and it will probably not be answered finally until we go to Mars and find out for ourselves.

It should be understood, however, that although the canals are a subject worthy of some attention, they are usually much less prominent and important as features than the large dark markings, as may be seen from trying to find some of the better-known canals on the accompanying photographs.

The satellites of Mars. — Early observers did not know of the Martian satellites, which are visible only in large telescopes. They were discovered by Asaph Hall, with the large refractor of Washington Observatory, at the opposition of 1877. The reason that they escaped detection for so long is that even through a powerful instrument they are very faint objects and their feeble light is usually swamped by the dazzling brilliance of Mars itself.

These two satellites were christened Phobos and Deimos (the Greek names for Fear and Terror, the attendants of Mars). Phobos revolves at a distance of 5,800 miles, and Deimos at 14,600 miles. These distances are between centres, however, so if we subtract the radius of Mars we find that Phobos travels at a height of 3,700 miles and Deimos at 12,500 miles above the surface of the planet.

The orbits of the two satellites are almost circular and both lie in much the same plane, within one or two degrees of the Martian equator. Being so near its primary, Phobos moves more rapidly than Deimos, making one complete revolution in 7h 39m, whilst Deimos, at a greater distance, takes 30h 17m.

We know nothing about the surface appearance or composition of the satellites, because even in a powerful telescope they appear merely as points of light. Their diameters have been estimated from their brightness, and seem to be about 10 miles for Phobos and 5 miles for Deimos. Thus they are not much larger than a sizable city on Earth, and they would both fit inside the bounds of London. A study of their variations in brightness in different parts of their orbit shows that, like the Moon, Phobos and Deimos always keep the same face turned towards their primary.

THE MINOR PLANETS

The minor planets or asteroids were unknown not only in ancient times but also for two centuries after the invention of the telescope. This does not mean that the early telescopes were not powerful enough: it is simply that the asteroids are all very small and mostly very faint, so that they become lost amongst the myriads of stars in the sky background. Indeed, as we saw in an earlier chapter, Piazzi discovered the first asteroid by chance, at Palermo, on 1 January 1801.

From this time on, the number of known asteroids began to increase: slowly at first, then more quickly; since the application of photography in the second half of the nineteenth century, the rate of discovery has continued to grow.

The discovery of the minor planets. — Shortly after the discovery of Ceres, Olbers discovered the second minor planet, Pallas, on 28 March 1802. On 1 March 1804, Harding found the third, Juno, and in 1807, Olbers found the fourth, Vesta. The latter is the brightest of the asteroids, since at favourable oppositions it is just visible with the naked eye. From this it appears that Vesta might have been discovered without the aid of a telescope by someone patient, thorough and keen-sighted enough.

It was not until 1845 that further members of this

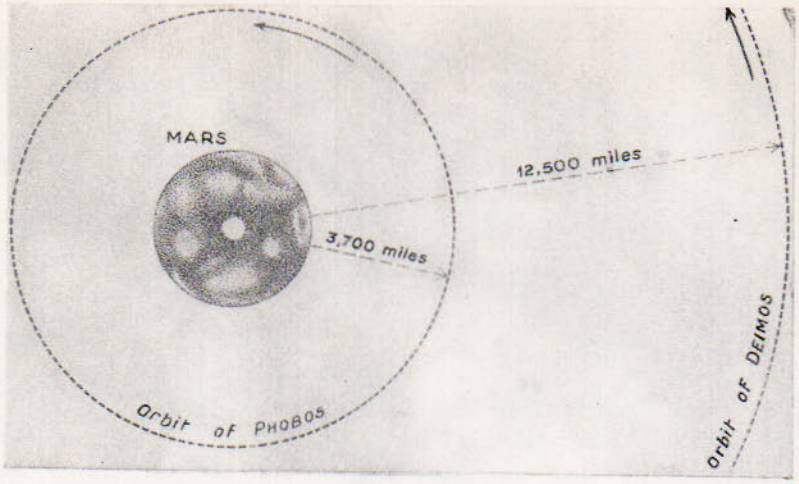


Fig. 399 — ORBITS OF THE MARTIAN SATELLITES

interesting family came to light, when Hencke found the fifth, to be followed by the sixth two years later. This stimulated new interest in the subject, and fresh asteroids were added to the list every year. As an example of the patience that was needed, an astronomer named Goldsmith, using a small refractor, discovered no less than fourteen between 1852 and 1861. The total number of known asteroids reached 100 in 1868, 200 in 1879, and 300 in 1890.

It was during this period that observational methods changed. Up to then, the procedure was to examine the sky through a telescope and compare what one saw with a star chart. If there was a 'star' in the sky which was not shown on the chart, it would be observed for several days running to detect any movement which would show it to be an asteroid. The patience and skill required by the observer were enormous.

In contrast to this, the 'labour-saving' photographic method was introduced by Max Wolf, the German astronomer, at Heidelberg Observatory, in 1892. This gave much more fruitful results.

Unlike a chart, a photograph rapidly gives the observer a reliable map of any portion of the sky in which he is interested, no matter how rich in stars that area may be. If an asteroid is in the field of view, it gives a different sort of image from a star. What happens is that the telescopic camera moves at the same rate as the stars, and so the stars, which do not move in relation to the plate, give point images. An asteroid, on the other hand, is moving in relation to the stars, just as the Moon is, and therefore leaves a trail-image on the plate, the length of the image depending on the length of the exposure and the speed of the asteroid's apparent motion. Another method is to examine two plates, which have each been taken with a short exposure at an interval of a couple of days, through a stereoscope. The stars will appear to be at infinity, but the asteroid, because of the change in its position between the two plates, will appear to stand out in front of the stars. This gives a very easy way of picking out a faint asteroid from a rich star-field.

By this method the task of locating asteroids was considerably simplified, with a consequent increase in the numbers discovered. At the present time about 2,000 of

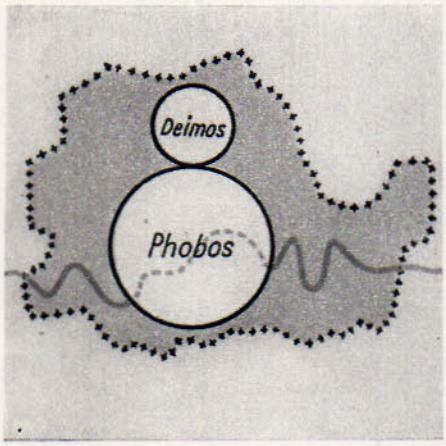


Fig. 400 — THE SIZES OF PHOBOS AND DEIMOS COMPARED WITH LONDON

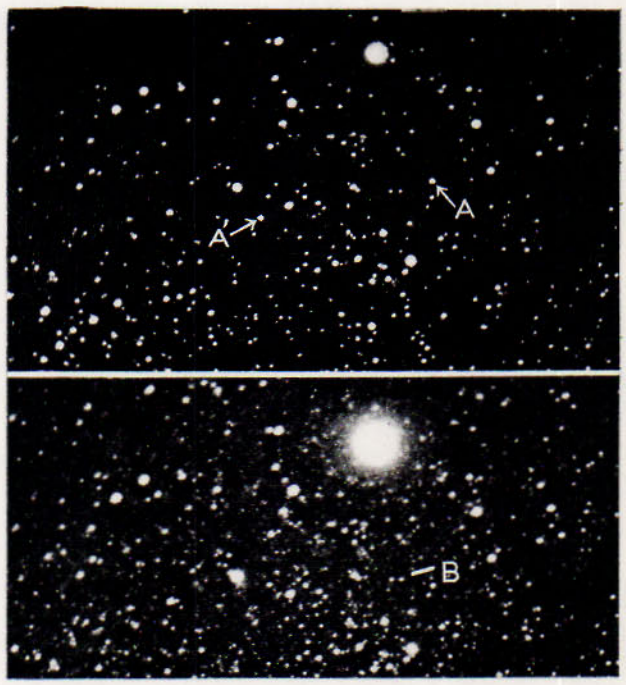


Fig. 401 — THE DISCOVERY OF A MINOR PLANET IN A STAR-FIELD. Above, visual method, by its movement between A and A' from one day to the next; below, photographic method, whereby the asteroid produces a trail owing to its movement during the exposure.

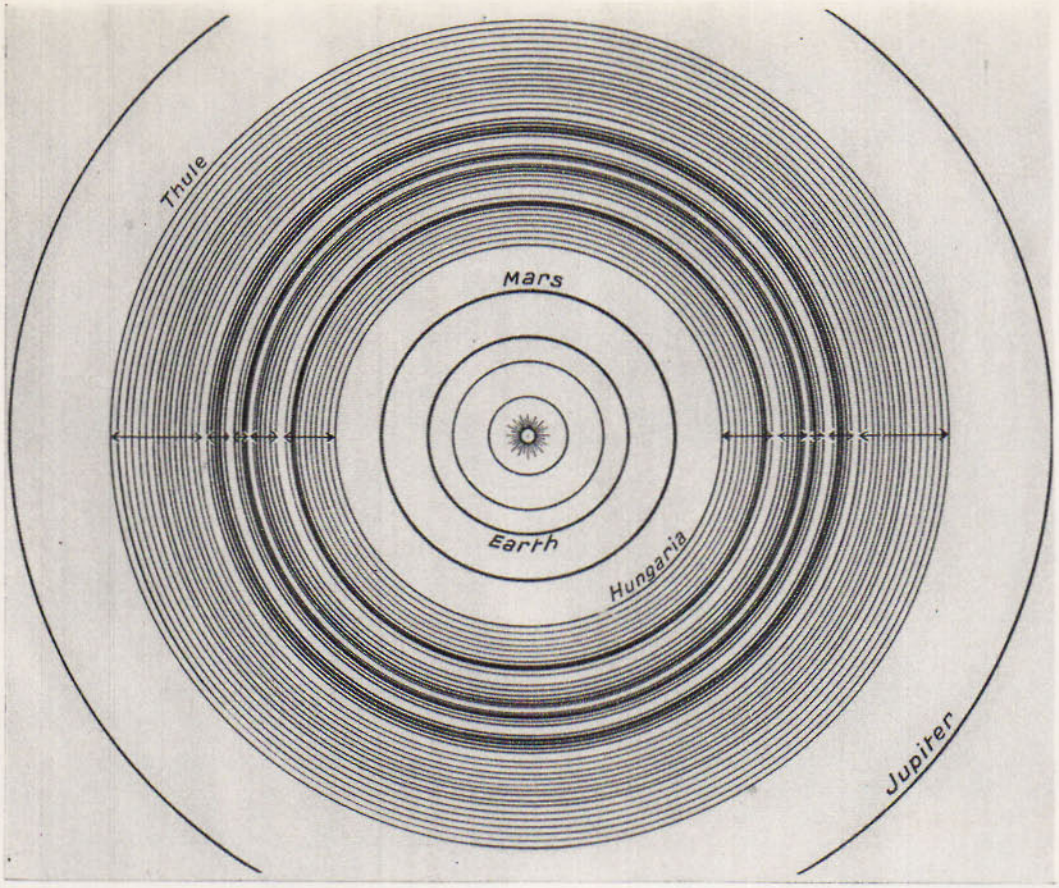


Fig. 402 — DIAGRAM OF THE DISTRIBUTION OF THE ORBITS OF THE MINOR PLANETS BETWEEN MARS AND JUPITER

these tiny bodies have been catalogued. We say 'about' because it is sometimes hard to tell, without some considerable mathematical labour, whether a 'new' discovery is really new or whether it is merely a previously known asteroid which became lost over the years. Nowadays, indeed, the old problem has been reversed: it is much more difficult to keep track of all the discoveries than actually to discover new asteroids. It involves computing an asteroid's orbit from all the available observations in order to predict where it will be in future. Without this knowledge there is a grave danger of a new asteroid being lost again, only to be rediscovered a few years afterwards and hailed as an entirely different body.

Distribution of the minor planets. — The great swarm of minor planets destroys the simple picture of the Solar System which most of us are taught at school. We know nowadays that instead of all being nicely packed between the orbits of Mars and Jupiter, forming a ring of asteroid dust, many of the minor planets wander far out of this zone. Some come right in towards the Earth, and Icarus approaches the Sun even nearer than Mercury does; on the other hand, some asteroids, e.g. Hidalgo, travel out as far as the orbit of Saturn. Furthermore, many asteroids do not revolve in the same general plane as the planets, their orbits being inclined by as much as 30° or 40° to the plane of the Solar System.

A complete treatment of the motions of the minor planets would require a volume of celestial mechanics to itself. Here, therefore, we shall attempt only to summarise the salient points.

Considering the distribution of the minor planets as a whole, it is clear that they lie mainly between the orbits of Mars and Jupiter. Their distribution within this zone, however, is far from being uniform. The combined gravitational attractions of certain of the major planets tend to concentrate the asteroids into a number of favoured positions, and so, for example, there are large numbers of them at 2.7 and 3.2 A.U. from the Sun. Thus a chart of these orbits seen in plan takes the form of a series of concentric zones varying in density.

On the other hand, there are a number of cases of remarkably high orbital eccentricity. The mean distance of Eros from the Sun is less than that of Mars, and other differences are even more pronounced. For example, at perihelion, Hidalgo lies a little beyond the orbit of Mars, but at aphelion it is not far from the orbit of Saturn. Apollo and Adonis come within the orbits of the Earth and Venus, and, as mentioned above, Icarus has the singular distinction of passing nearer the Sun than Mercury.

In 1937, however, Reinmuth discovered Hermes, which passed within 400,000 miles of the Earth. Cunningham calculated that the orbit of Hermes is the smallest of all the minor planets, and it almost exactly crosses the Earth's path. This means that at certain times Hermes could pass between the Moon and Earth at a distance of about 221,000 miles. It must be noted, however, that there is still some uncertainty attached to this figure, since owing to insufficient data the orbit of Hermes has been only provisionally determined.

Thus it can be seen that to both the theorist and the specialist in celestial dynamics, the asteroids are objects of considerable interest on account of their behaviour

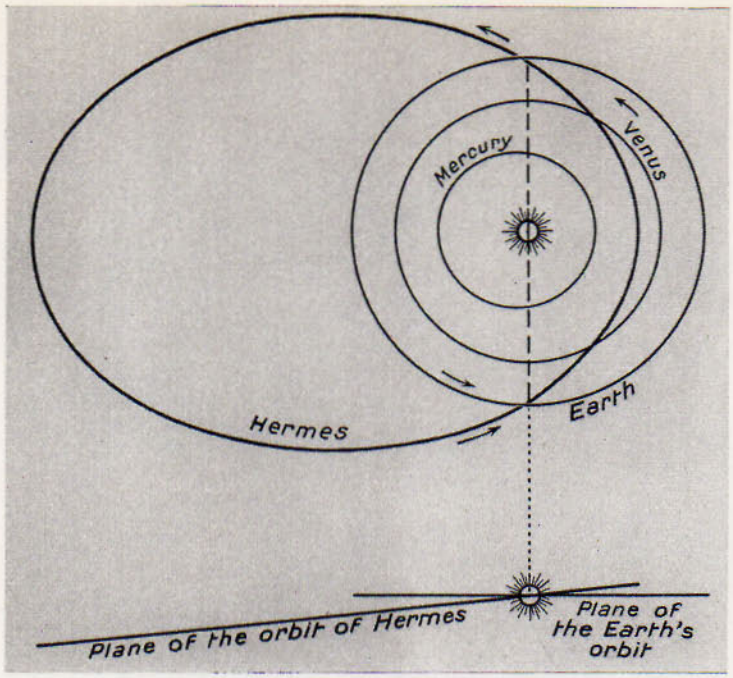


Fig. 403 — ORBIT OF THE PLANET HERMES

and distribution in the Solar System. Indeed the question of their origin and their exact relationship to the other members of the Sun's family has not yet been satisfactorily answered.

Size and nature of the Asteroids. — In the main, the diameters of the minor planets cannot be determined directly except with very powerful instruments; even so, the largest asteroids appear only as tiny dots, and are extremely difficult to measure exactly.

Measurements have in fact been attempted, but they are subject to the possibility of error, as indicated by the disagreement between various authorities. The following are the most recent values for the diameter of the 4 largest asteroids:

Ceres	427 miles	Pallas	280 miles
Vesta	240 miles	Juno	150 miles

Provisional values have also been obtained for the following: Hygeia, Eunomia and Davida, 213 miles; Hebe and Iris, 176 miles; Psyche and Patroclus, 169 miles; Laetitia and Dembowska, 161 miles.

Although these values cannot be accepted absolutely, they give some idea of the approximate size of these asteroids. Compared with the Moon, for example, they are minute.

The others are even smaller and any knowledge we may have about them is based on photometric studies. For these purposes, we assume that their surfaces are much the same as any other planet's (excepting the gaseous giants), in which case the amount of light they reflect is proportional to the surface area. Combining their apparent diameters with their distance, it can be concluded that most of the minor planets are only a few miles across, and the smallest yet found are not much more than 1 or 2 miles in diameter at the most.

Thus it is difficult to dignify these minute bodies with the name of 'worlds' in the proper sense of the word. Furthermore, we do not even know if they are spherical, although the largest, shown in Fig. 404, are this shape. Various facts indicate that sphericity is not universal among asteroids.

Many asteroids, indeed, show continuous variations in brightness which may be explained on the hypothesis that they are rotating about an axis (of any inclination) in such a way as to present different surface areas to the Earth, thus changing the amount of reflected light. It is further supposed that many of the asteroids are not even regular

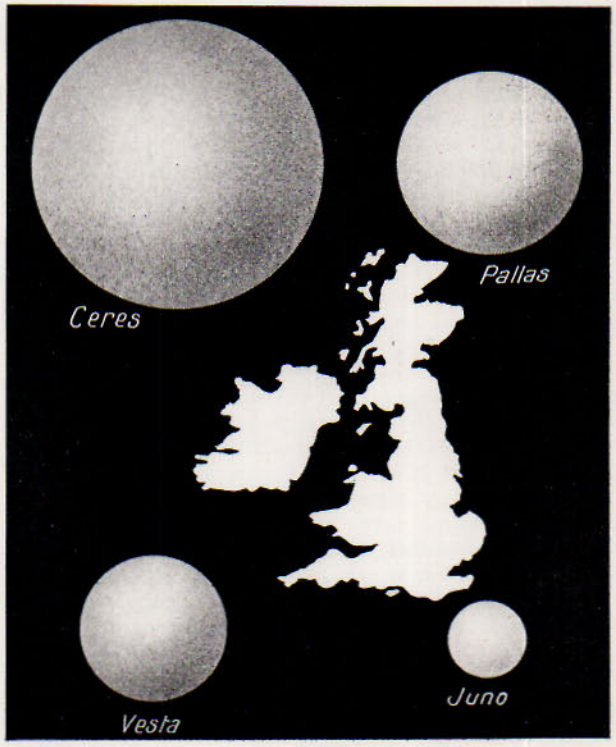


Fig. 404 — SIZES OF THE FOUR LARGEST ASTEROIDS COMPARED WITH THE BRITISH ISLES

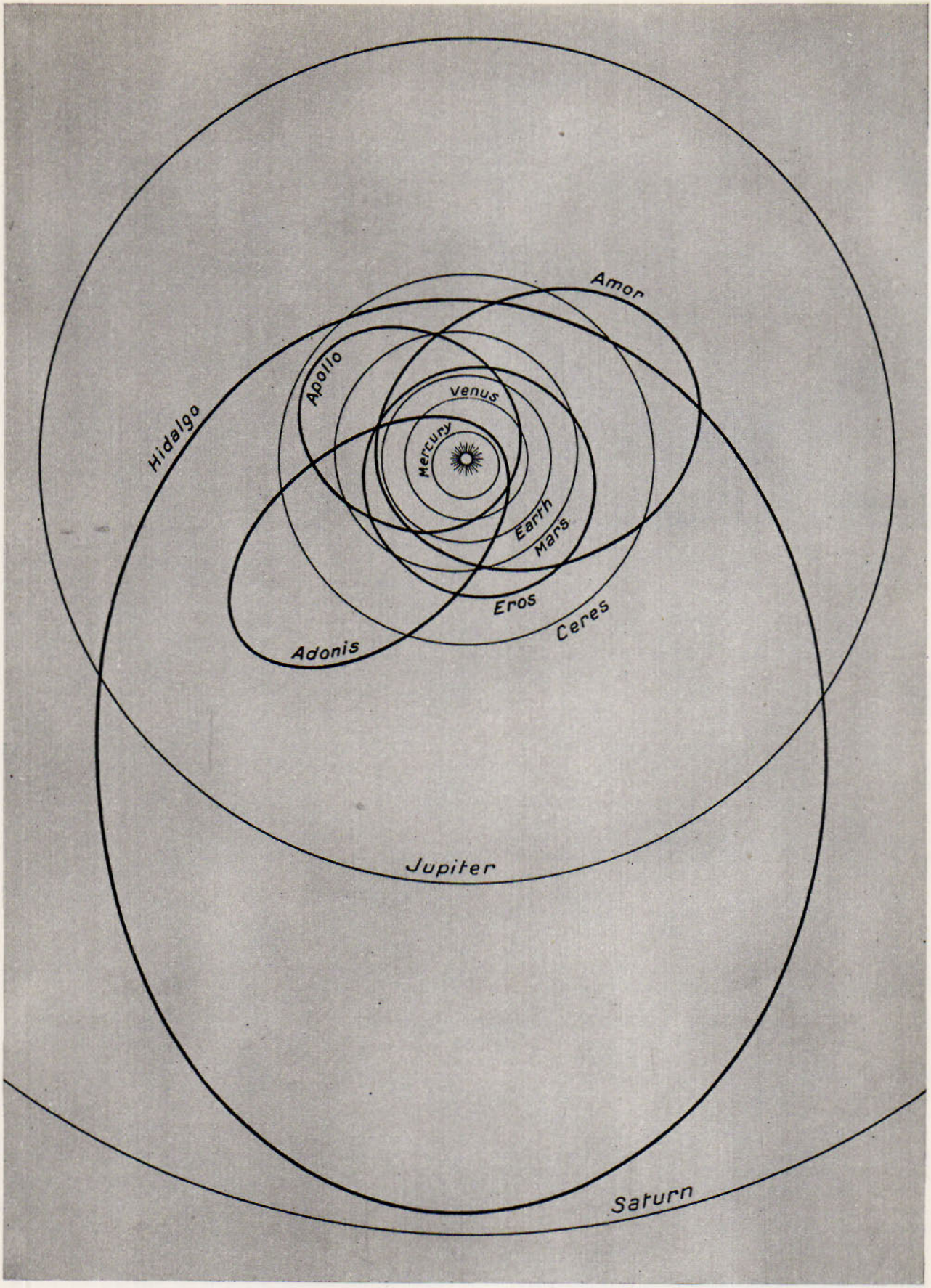


Fig. 405 — Some eccentric asteroid orbits

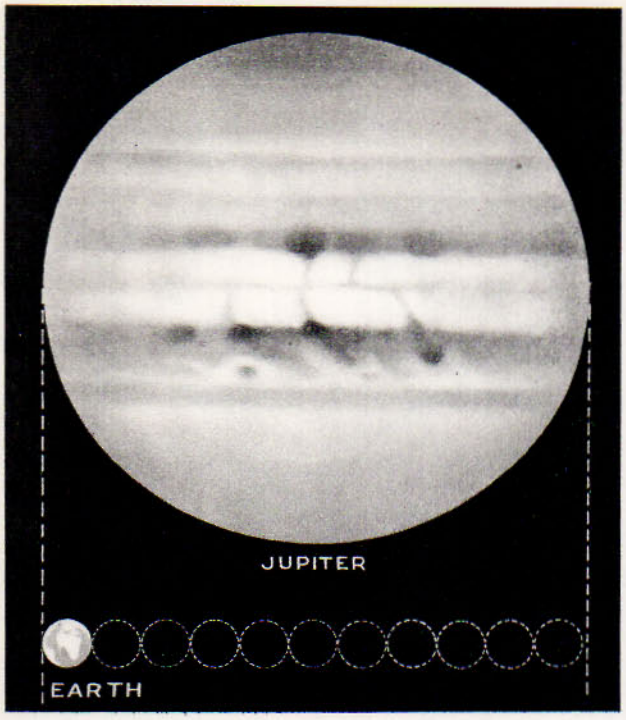


Fig. 406 — THE SIZE OF JUPITER COMPARED WITH THE EARTH

in shape, but are simply huge blocks of material with angular faces. The case of Eros is even more curious; its variations are such that a number of different hypotheses have been proposed to explain them — perhaps a pair of small bodies associated to form a 'double planet'; perhaps a very elongated block of material rotating about its smallest diameter; or perhaps a crystalline structure which gives a different amount of reflection as the Sun strikes the facets at different angles. The second suggestion seems the most plausible, but the matter is still open.

Thus we still have much to learn about the nature of these minute bodies, and this will require a great deal of work with powerful instruments, using refined techniques. We do not know for certain, for example, whether any of the asteroids possess an atmosphere, but from theoretical considerations alone the answer is probably in the negative. They are so small in mass that it is inconceivable that any of them could have retained an atmosphere.

The last question we have to consider is the nature and composition of these minor planets, which form a sort of ring of dust, interspersed with larger fragments, round the Sun. To answer this question, we must know how the asteroids originated, which is a very controversial matter. At all events, it would seem that they have a common origin. The main hypotheses are (a) that the asteroids were formed from material which, when the Solar System itself was formed, could not condense into one large planet owing to perturbing influences, or (b) that they are the broken fragments of a primitive planet which was disrupted through collision or excessively rapid rotation at some early date. The second suggestion would go to explain the suspected irregularities in the shape of the asteroids. A final suggestion is that the asteroids are the nuclei of ancient comets.

Whatever the truth of the matter, the minor planets as a whole represent only a very small amount of material; their total volume does not exceed a thousandth that of the Earth, which would make the hypothetical proto-planet very small indeed.

JUPITER

Dimensions of the globe. — Beside the other members of the Solar System, Jupiter is a colossus: its mean diameter is 85,750 miles, i.e. 10.08 times the size of the Earth's. Jupiter's volume, moreover, is 1,312 times as great as the Earth's. But this enormous globe is surprisingly light: its mass is only 314 times the Earth's, so that its density is only 1·34.

Even from the early days, it was realised that conditions on this vast, relatively light planet must be very different from those on Earth, and this is supported by Jupiter's surface appearance.

In the first place, even a cursory glance through a small telescope shows a most pronounced flattening of the Jovian disc, the polar diameter being almost one-fifteenth shorter than the equatorial. It can be shown that this ellipsoidal form is due to the great centrifugal force developed at the planet's equator by the extremely rapid rotation.

The rotation period of Jupiter is the shortest in the Solar System, being only about 9h 58m. On the other hand, there are no great variations in the length of Jupiter's short day and night, as there are for the other planets: the inclination of its axis is only 3°, and seasonal effects are consequently very insignificant.

From these data, it can be calculated that the force of gravity on this enormous world (neglecting the effects of the high rate of spin) is 2·53 g, which means that a 12-stone space-traveller would have the unpleasant experience of continuously weighing 30.5 stone if he landed on Jupiter. After a time he would probably collapse under his own weight, provided that there was a solid surface to support him. This point will be discussed later.

Orbit and motion of Jupiter. — The eccentricity of Jupiter's orbit (0.0484) means that the planet's distance from the Sun varies considerably: at perihelion it is 459,800,000 miles, and at aphelion 506,800,000 miles. When Jupiter is closest to the Earth (at a perihelic opposition) the two bodies are separated by a distance of about 365,000,000 miles.

The size of Jupiter's orbit is therefore considerable, and the planet takes 11y 315d, at a speed of just over 8 m.p.s., to make one complete revolution. This means that Jupiter's movement against the stars is not very rapid, and it stays in the same region of the ecliptic for quite a long time. Jupiter returns to opposition about

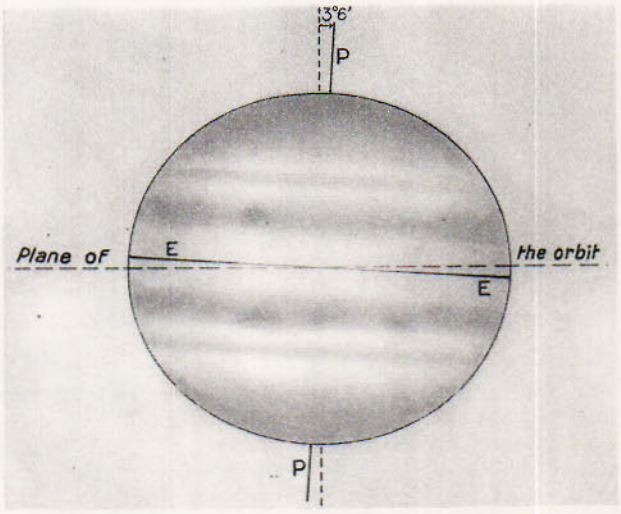


Fig. 407 — INCLINATION OF JUPITER'S AXIS OF ROTATION

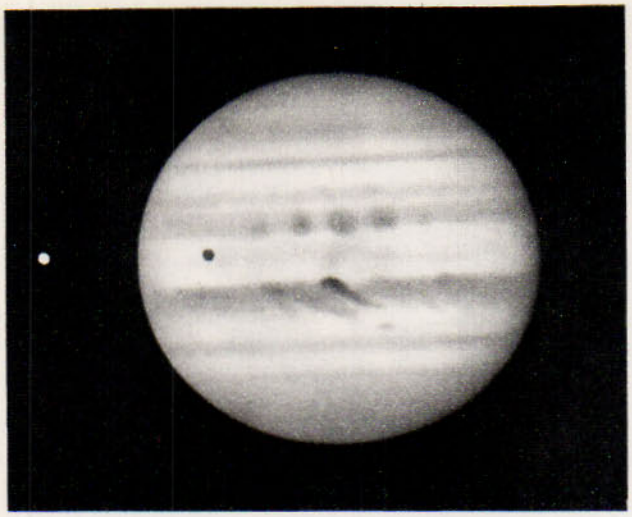


Fig. 408 — THE PHASE-EFFECT ON JUPITER, revealed at the quadratures by a slight darkening along the limb nearest the terminator.

a month later every year, its synodic period being 398·9 days.

Jupiter is so far away that, relatively speaking, the Earth does not move very far to either side of the line joining Jupiter and the Sun, and therefore Jupiter always presents a fully illuminated disc to us. When the Earth is furthest to one side of this line, a slight darkening along one limb of the planet marks the theoretical phase-effect.

The relative sizes of the two orbits are such that the variations in the separation of Jupiter and the Earth are very small. Thus the apparent diameter of the Jovian disc is 48″ at opposition and 30″ at superior conjunction.

Appearance of Jupiter. — Because of its great size, Jupiter is a brilliant object when seen from the Earth. With the naked eye it looks like a very bright star, one of the most impressive in the sky: its only rivals are Venus, and Mars when at perihelic opposition. When Jupiter is most favourably situated, its light is strong enough to shine through moderately thick clouds and to cast a shadow on a white surface.

After Venus, which can be seen as a crescent through a pair of binoculars, Jupiter is the easiest planet to recognise as such. At opposition its disc is only 40 times smaller than the Moon in diameter, which means that a very low magnifying power will show it as quite a sizable object. A strong pair of field glasses will reveal a small dot, whilst a naval telescope will clearly show the polar flattening.

The most characteristic feature of Jupiter is its system of belts encircling the globe parallel to the equator. They are sufficiently prominent to be seen in the most modest of astronomical instruments, and have been known from almost the earliest days of the telescope. The detailed structure of the belts is not permanent, and, as will be shown later, consists of atmospheric features, not surface markings of the type found on Mars or the Moon.

At first sight, especially if the image is on the small side, Jupiter's belts, with their different widths and intensities of shading, look very irregular. Separated by zones of varying brightness, they extend to a certain latitude either side of the equator only, the surface around the polar caps being of a uniform grey colour. All round the edge of the disc there is a slight darkening due to light-absorption.

Neglecting the rapid changes occurring within the belts and the fact that a whole belt occasionally disappears completely, the following nomenclature has been adopted. The equator is usually occupied by a bright zone, called the Equatorial Zone. On either side of this are the North and South Tropical Belts, these being usually the most prominent. Then come the North and South Temperate Belts, and finally the 'Polar Caps'. There are further subdivisions in each of these. In all the belts, especially the tropical ones, variations of colour occur; these range from pink to purplish brown. The 'polar caps', however, are of an indeterminate grey, sometimes with a trace of green or yellow.

Superimposed on this general picture are a number of smaller features. With high magnification on a large telescope it is possible to see all kinds of dark markings in the belts, some more or less circular, others elongated or irregular in shape. These spots are scattered along the belts or their edges, which gives the belts a rather 'knobbly' appearance. Occasional dark spots make their appearance in the bright zones, in which small, very white spots are also frequently visible. Dark streaks, either straight or curved, are also found, sometimes linking two belts but more often lying at an angle across a single belt.

All this forms a definite system of great complexity, and the swiftly changing spots and belts give the impression of being vast cloud-formations, violently churning against and over one another. We do not know precisely how this system was formed, nor what its underlying causes are, but the variations in Jupiter's appearance make it quite certain that we are not seeing the solid surface of the planet. In this respect Jupiter is more like Venus than Mars or the Moon (although from its density we know that Venus must have a solid surface beneath the clouds).

To obtain some idea of the changes, both large and small, in the appearance of Jupiter, it is sufficient to glance at a series of drawings or photographs spread over a few years. As will be seen from the illustrations in the text, the width and intensity of most of the belts are constantly altering: a belt will frequently split into two, or even vanish completely for a while. Much the same thing occurs in the case of the light and dark spots and the irregular streaks. Summing up the first impression, one might say that Jupiter seems to go through periods of violent activity, when the disc is full of small features suggesting extreme turbulence, to be followed by periods of comparative quiescence, when there is not much to be seen. It is uncertain whether there is any periodicity in these changes, which occur according to laws of which we are as yet ignorant.

In the midst of all this turmoil, however, one feature can be picked out. It was first observed by Cassini in 1665 and has been visible on and off ever since. Apparently permanent but extremely mysterious, the Great Red Spot is about 30,000 miles long and 7,000 miles wide, and in 1878, when it was most prominent, was of a bright red colour. It is situated in the South Temperate Zone, and is usually confined to the temperate belt; when the South Tropical Belt is present, however, the Spot appears to divert the edge of the Belt around itself to form a vast, white, egg-shaped 'bay'.

This 'bay' can be seen in various drawings and pictures in the text.

The most mysterious thing about the Great Red Spot is that it gradually drifts in longitude, like an island floating on a sea. Careful observations over a period of fifty years show that the Spot has drifted in longitude in the following way: taking its 1910 longitude as 0°, in 1911, its longitude was slightly variable around 325°; in 1912,

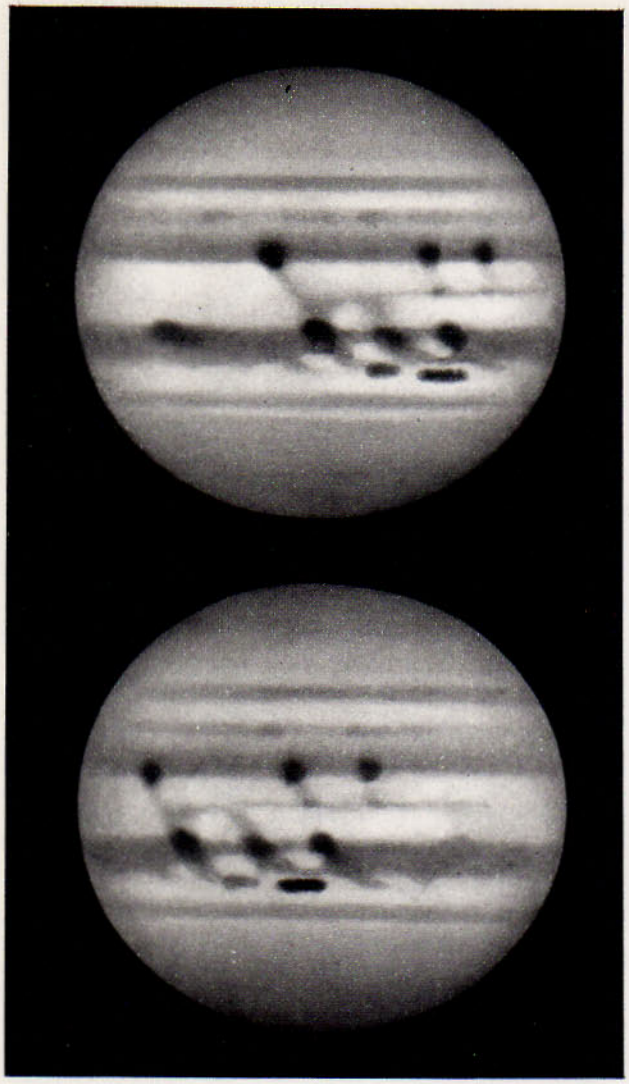


Fig. 409 — JUPITER'S RAPID ROTATION. Two views of the planet, taken one hour apart.

it reached 303°; in 1913, 260°; in 1914, 200°. During this time, therefore, the Spot moved through a distance of almost 44,000 miles, which is considerably more than its own length. Again, in 1937, its longitude was 138°; in 1939, 170°; in 1940, 176°, but in 1941 it returned to 170°. Another interesting feature is that its colour and also its visibility change at the same time. There are times when the Spot does not live up to its name, becoming a greenish-white colour and hardly visible. Then there will suddenly be a recrudescence of colour which makes the Spot quite prominent in even a small telescope. At such times it is of a very pronounced salmon-pink colour.

The curious behaviour of the Red Spot appears to bear some relation to the behaviour of the Tropical Belt; the Spot is of a deeper colour, and therefore more easily visible when this belt becomes faint or almost invisible.

What the causes of this behaviour are, no one can tell, and we are left simply with the mute facts. Equally puzzling are the drift in longitude, and the fact that the Red Spot is an apparently stable formation among all these other spots and streaks. Indeed, the behaviour of the surface features generally is puzzling: they seem to slide past each other unaffected, or undergo some mutual attraction which slightly deforms them, or, once in contact, drag each other along. This is exemplified by another striking feature, the South Tropical Disturbance. It first became visible at about the beginning of this century, but by 1918 its extension in longitude was such that it embraced almost half the planet. Its period of rotation was shorter than that of the Great Red Spot, which meant that there was a relative movement between them. Whenever the Disturbance overtook the Spot, there was a violent interaction, and the latter was dragged round with the Disturbance for a while before they drifted apart again.

This emphasises the fact that the rotation speeds at different latitudes on Jupiter are not the same. The surface may be divided into a number of zones with different periods: 9h 50m for the equatorial regions, and 9h 55.3m for the regions on either side. This means that we are not dealing with a solid globe, turning as a whole, as in the case of Mars or the Moon.

The composition and physical condition of Jupiter. — The details described in the preceding section, particularly the mottled aspect and the instability of the belts and smaller features, indicates that we are dealing with a very different type of planet from those we have encountered up to now, and it is very difficult to come to any conslusions as to Jupiter's exact nature. It seems that Jupiter is composed of a substance which behaves like a fluid in certain respects: picture it as a huge sphere of gas at such high pressure through its own gravity that it has become almost liquid. This would explain the pronounced flattening of the disc by the centrifugal force of its rapid rotation.

At all events, it can be said with confidence that the appearance is due simply to our seeing the outer layers of an immensely thick atmosphere, which hides whatever lies below. There is probably no 'surface' in the proper sense of the word. Spectroscopic analysis shows that the atmosphere contains large quantities of methane and ammonia, and it is probable that hydrogen and helium are also present. Much, however, still remains to be learned about the exact nature of Jupiter's atmosphere and the origin of the turbulent exterior features which are observable from the Earth.

It is now known that, contrary to what was once thought, Jupiter is not a world in process of formation, i.e. giving out light and heat from its own internal resources. This can be proved by studying its satellites. When one of these moons passes between Jupiter and the

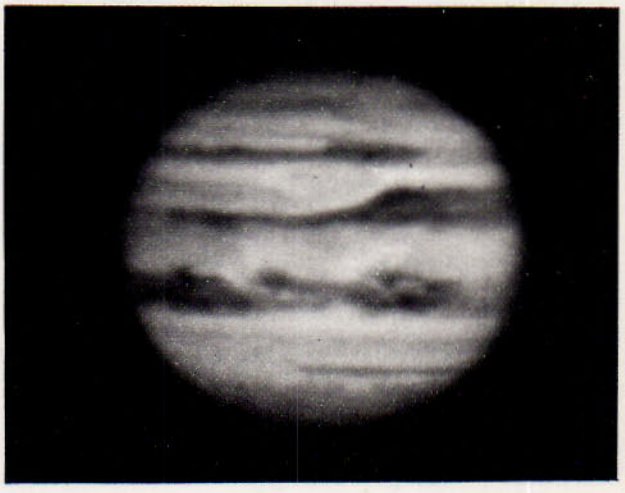


Fig. 410 — JUPITER AT 21h 45m ON 20 MARCH 1945. A composite image from photographs taken by H. Camichel through the 24-inch refractor of the Pic du Midi Observatory.

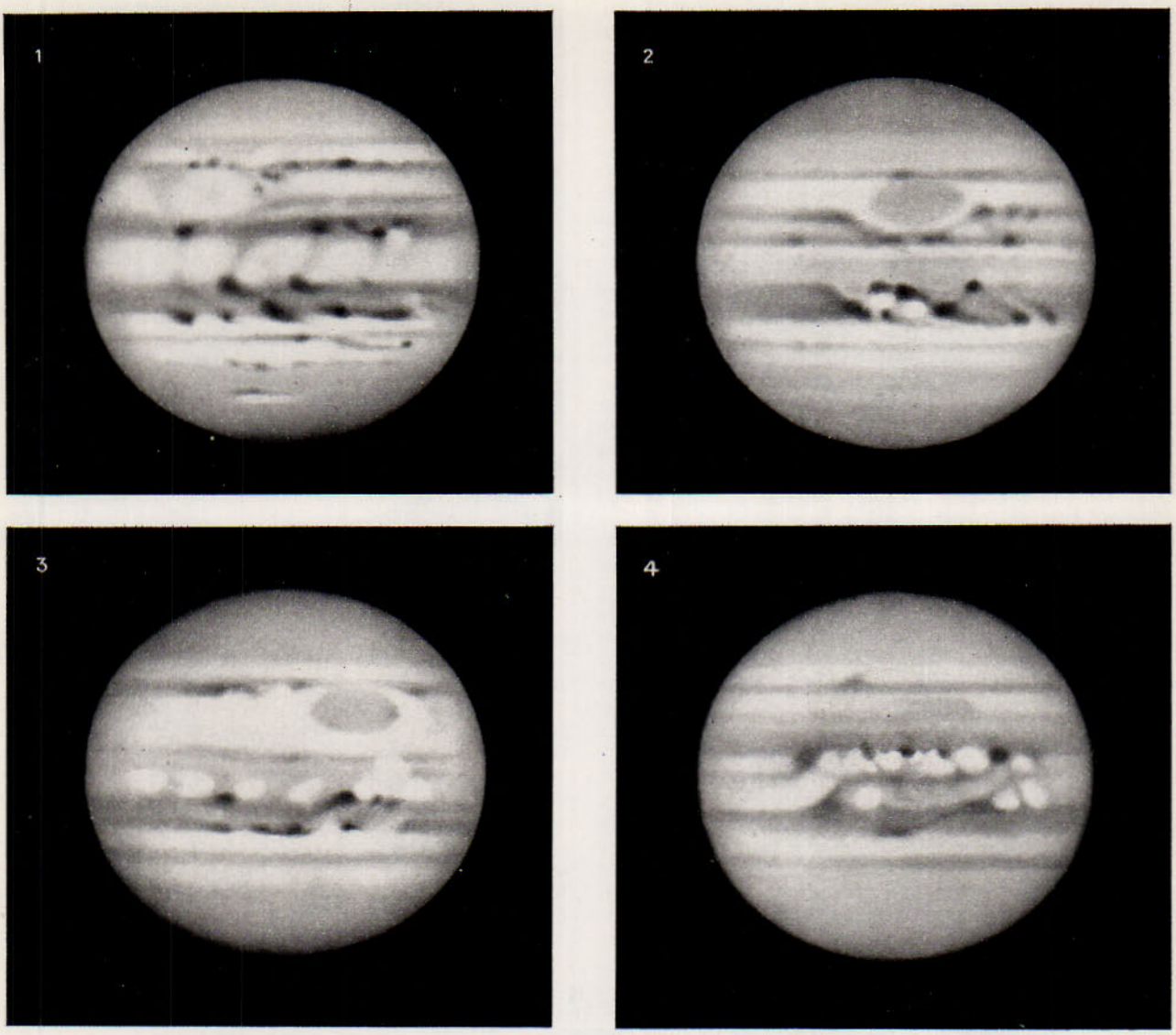


Fig. 411 — DIFFERENT ASPECTS OF JUPITER. 1, 1929; 2, 1933; 3, 1936; 4, 1938. *(Drawings by L. Rudaux)*

Sun, its shadow is perfectly black against the surface of the planet, which is brilliantly illuminated by the Sun. When a satellite passes through Jupiter's shadow, it completely disappears, whereas if Jupiter were intrinsically luminous it should be very faintly visible during the eclipse. Moreover, measurements made with thermocouples indicate a very low temperature on the planet, in the region of $-135°$ to $-155°$ C.

This immediately puts out of court the old notion that Jupiter is a world more like the Sun than the Earth, thus being a stellar rather than a planetary body. No more probable is the picture of Jupiter as a world very similar to our own, but simply covered in much denser layers of cloud.

The most recent theories picture Jupiter as consisting of a large inner core of ice (under enormous pressure, of course) surrounded by a vast ocean of compressed and liquefied gases, merging into an atmosphere of hydrogen. In this outer atmosphere, clouds of methane and ammonia, violently churned by the differential rotation, form the belts and spots visible from the Earth.

The satellites of Jupiter. — At present there are 12 known satellites of Jupiter; only 4 of these, however, are of any great size, and they were discovered by Galileo as soon as he turned his newly made telescope on the planet. Whilst the Galilean satellites are visible through even a small telescope, the remainder can be detected only with powerful instruments, preferably employing the aid of photography.

In the table showing the features of these satellites, on page 216, the numbers given to them probably look rather peculiar at first glance. The reason is that the Galilean satellites were numbered I, II, III, IV, in order of distance from Jupiter, whilst the others were numbered in the order of their discovery.

The distances of the outer satellites from Jupiter itself show how vast the system really is, and there are quite a few interesting details which are worth mentioning. The first 5 satellites, whose orbits are almost circular, revolve in more or less the same plane, which practically coincides with the plane of Jupiter's equator. The orbits of the rest, however, are of high eccentricity and are steeply inclined to the plane of the Jovian equator. Moreover, the movement of satellites VIII, IX, XI and XII is retrograde.

Satellites I to IV are the most interesting because they are the easiest to observe. Two phenomena can be studied

to advantage through a medium-sized telescope: transits of a satellite acros Jupiter's disc, on to which the satellite frequently casts a shadow when suitably placed; and occultations, when the satellite disappears behind the disc, or eclipses, when the satellite disappears in Jupiter's vast cone of shadow.

The conditions under which these phenomena occur vary according to the positions of the Earth and Jupiter relative to the Sun. Fig. 415 shows the principle involved, and also the changing effects of perspective due to the relative motion of the three bodies. Thus at opposition the line joining Jupiter and the Earth is coincident with the direction to the Sun: when a satellite passes in front of the planet's disc, it is also possible to see the shadow it casts. At other times, particularly at the quadratures, the Earth is at an angle to the direction of illumination, and the shadow of the satellite appears to be some little way behind or in front of the satellite itself. It then quite frequently occurs that a satellite can be seen passing in front of the disc unaccompanied by a shadow, or the shadow is still well on the disc before or after the satellite transits. For the same reason, a satellite passing behind Jupiter may disappear into the cone of shadow before actually being occulted by the limb of the planet. At opposition, this cone of shadow is exactly hidden by the globe, and the satellites are occulted without first being eclipsed. Indeed, there is a whole series of interesting phenomena to be seen when Jupiter's shadow cone is projecting back at an angle to our line of sight. For example, a satellite will disappear behind one limb of the planet; then, during the occultation, it will plunge into the shadow cone, and only become visible again when it emerges some distance to one side of the disc.

Strictly speaking, all this applies only to Io, which is the closest to Jupiter. In the case of the others, there are a number of additional details. Because of the obliquity of their orbits, it is sometimes possible to see one or other of the Galilean satellites disappear behind the disc of Jupiter, reappear, travel a little further and then vanish into the shadow cone.

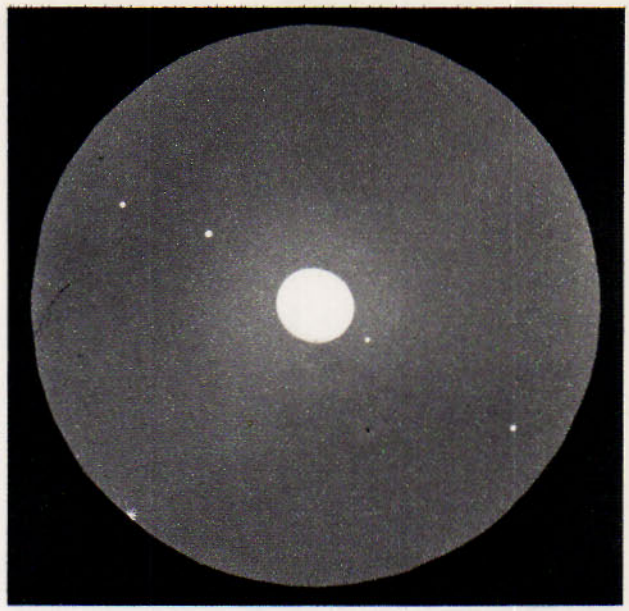


Fig. 412 — JUPITER AND ITS SATELLITES IN THE FIELD OF A SMALL TELESCOPE

These transits, eclipses and occultations are repeated in the same order at every revolution of the three inner satellites. The reason why this does not apply to the fourth one is simple. The satellites revolve very nearly in the plane of Jupiter's equator, which is only very slightly inclined in relation to the Earth. Thus we can imagine their orbits as circles seen nearly or exactly edgewise on: sometimes we see nothing more than a straight line projecting either side of the planet; at other times, when the circles are slightly tilted, we can imagine them as very long, thin ellipses. Now the minor axes of the ellipses formed by the orbits of the first three satellites are shorter

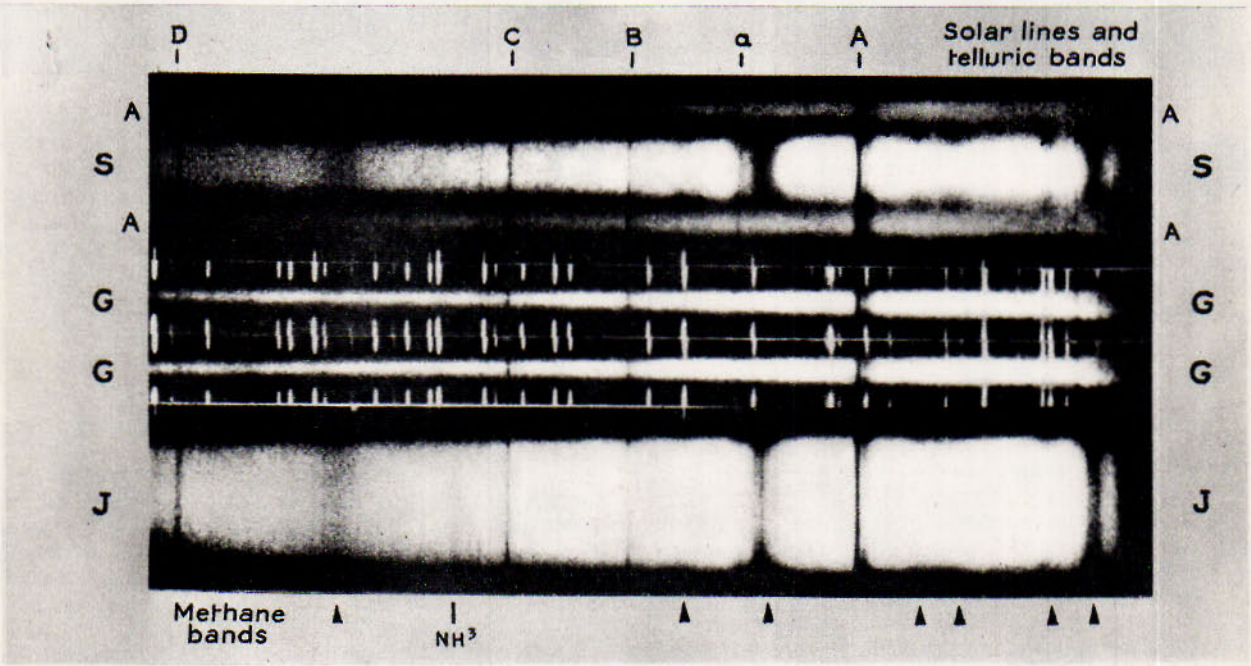


Fig. 413 — THE RED ENDS OF THE SPECTRA OF JUPITER, SATURN AND GANYMEDE, showing the bands of methane on Jupiter and Saturn, a band of ammonia on Jupiter, and only the solar and telluric lines on Ganymede and Saturn's rings. *(G. P. Kuiper, MacDonald Observatory, 1944)*

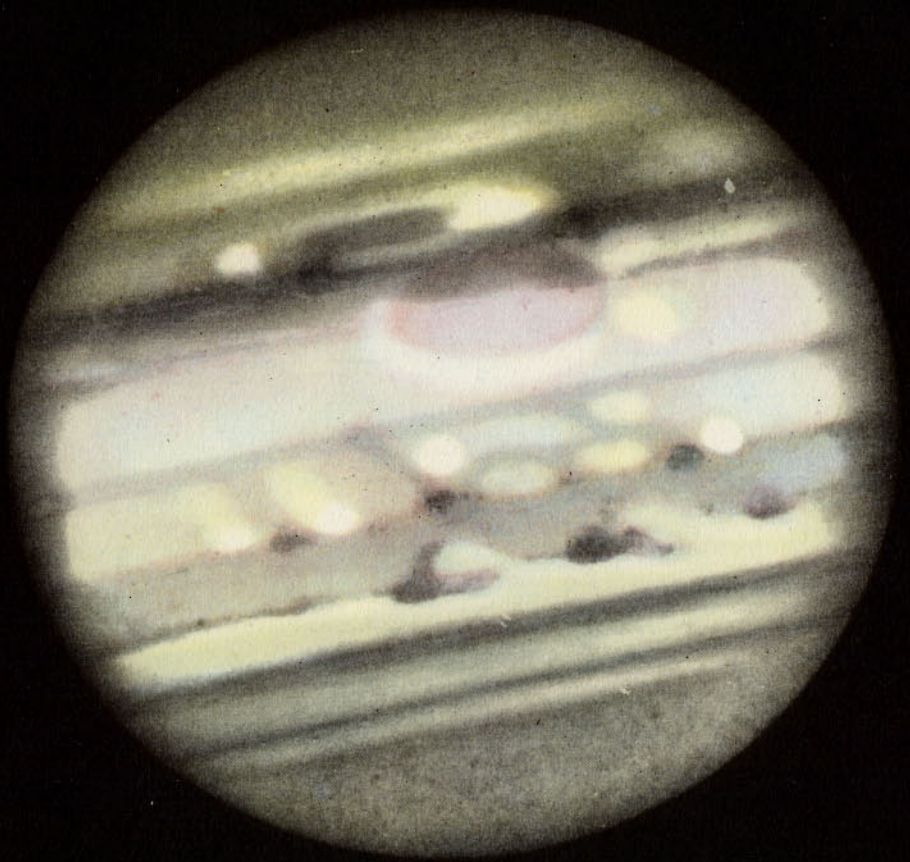


Plate IV. THE PLANET JUPITER

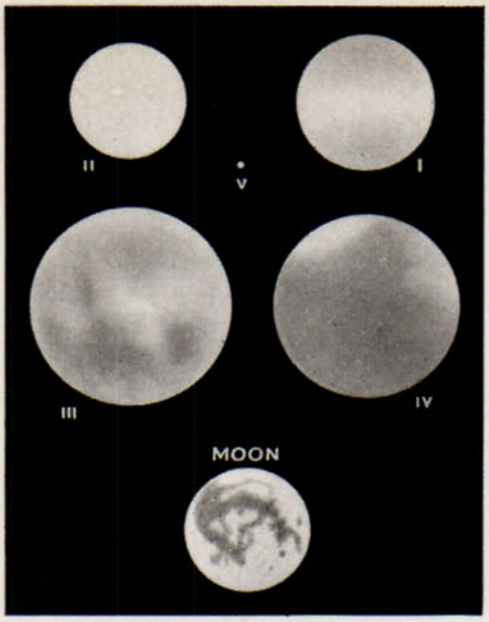


Fig. 414 — SIZES OF THE PRINCIPAL SATELLITES OF JUPITER COMPARED WITH THE MOON

than the polar diameter of Jupiter: consequently, at every revolution each satellite will pass once behind and once in front of the planet's disc, in opposite directions. But satellite IV (Callisto) has a more highly inclined orbit, whose minor axis, when seen from the Earth, is always longer than Jupiter's polar diamater. In this case, therefore, the satellite will pass either above or below the disc, but not in front of or behind it. Nor will there be an eclipse, for it misses the shadow cone too.

The spectacle of Jupiter's moons is interesting both in its own right and as an object-lesson in astronomy for the beginner with a small telescope. While on this subject, indeed, it may not be out of place to add a few remarks on the physical nature of Jupiter's satellites.

The figures already given show that the largest of them surpass the Moon in size, and even rival some of the planets themselves. Ganymede and Callisto are somewhere between Mercury and Mars in diameter, and in even quite a modest telescope they are visible as discs with a definite diameter.

When the satellites transit across Jupiter, they can usually be seen quite clearly as tiny light or dark discs (depending on the colour of the background). They do, however, show differences which are intrinsic to the satellites themselves, since they do not all have the same reflecting properties. This can be shown by comparing them with the surface of Jupiter when they are in transit; taking the same region as a background, one satellite may be of exactly the same brightness as the planet and will consequently be very difficult to see at all, whilst others will stand out as light or dark spots. Moreover, these contrasts are not uniform and it occasionally happens that only a part of a satellite is visible, the rest merging with the background. This proves in an elementary fashion that the surfaces of the satellites must be mottled with patches of different reflectivity. In large telescopes these patches can even be glimpsed, but it is impossible to be sure of their exact outlines.

Allowing for differences in optical power and the difficulties of the observational problem involved, it can be concluded from the work that has already been done that these markings are permanent features of the surface structure and, as such, they enable us to determine the rotation period of the satellite in question. It has been found that, like the Moon, they revolve keeping the same face to their primary, and this confirms what might be predicted from theory.

As to the question of whether any of Jupiter's satellites possess an atmosphere, there is no indication of gaseous matter in detectable quantities around any of them, and Kuiper, using spectroscopic methods, has confirmed this. However, the kinetic theory of gases would allow for the possibility that Ganymede at least might possess a very tenuous atmosphere.

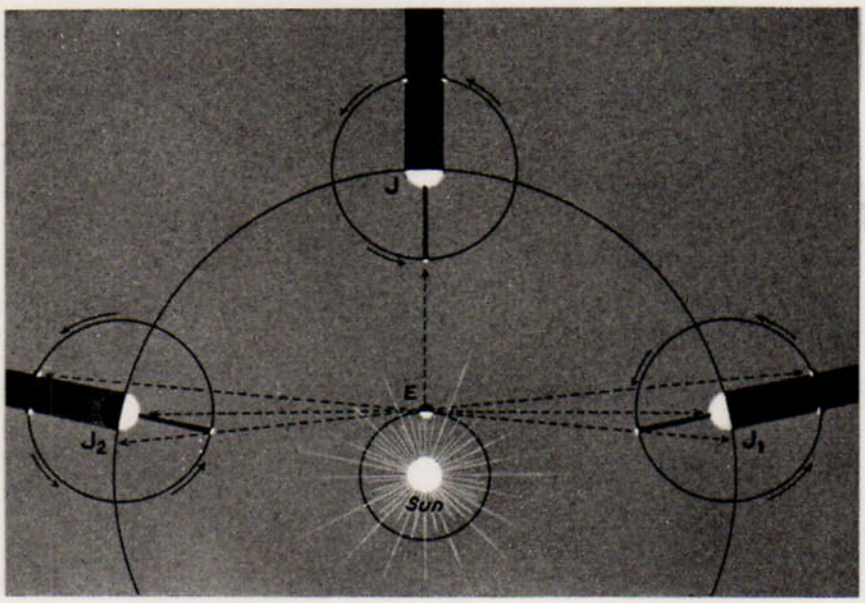


Fig. 415 — THE VARIOUS WAYS IN WHICH JUPITER AND ITS SYSTEM ARE PRESENTED TO THE EARTH. This governs the observation of the various phenomena connected with the motions of the satellites. E, Earth; J, J_1, J_2, Jupiter.

NUMBER	NAME	DIAMETER MILES	DISTANCE (10^3 MILES)	REVOLUTION PERIOD	INCLINATION OF ORBIT OF PRIMARY	DISCOVERY
V.	Amalthea	100?	113	0d 11h 50m	3°	Barnard 1892
I.	Io	2,310	262	1d 18h 27m	3° 7′	Galileo 1610
II.	Europa ...	1,750	417	3d 14h 13m	3° 6′	„ 1610
III.	Ganymede	3,200	666	7d 3h 42m	3° 2′	„ 1610
IV.	Callisto ..	3,220	1,170	16d 16h 32m	2°43′	„ 1610
VI.	(Hestia) ..	100?	7,120	250d 16h	28°45′	Perrine 1904
VII.	(Hera) ...	30?	7,290	259d 16h	27°58′	„ 1905
X.	(Demeter)	25?	7,300	260d 12h	28°24′	Nicholson 1938
XII.	(Adrastea)	15?	13,000	625d	160°	„ 1951
XI.	(Pan)	20?	14,000	692d	163°37′	„ 1938
VIII.	(Poseidon);	35?	14,600	739d	148° 4′	Melotte 1908
IX.	(Hades) ..	20?	14,700	758d	156°	Nicholson 1914

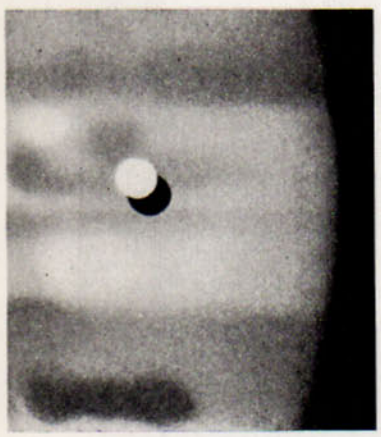
NOTE: The names in brackets are not yet official. There is considerable uncertainty in the diameters and revolution periods of Jupiter's outer satellites.

SATELLITES OF JUPITER

SATURN

Dimensions of the globe. — After Jupiter, Saturn is the second largest planet in the Solar System, and it can be classed among the true giants.

With a diameter of 75,100 miles, Saturn is 9.5 times larger than the Earth and 763 times greater in volume. Luke Jupiter, Saturn is of very low density, and the globe is strongly flattened, in the proportion of 1 to 10. Here again, the reason lies in the centrifugal force generated by the extremely rapid spin. Saturn's rotation period is about 10h 15m. The material out of which the planet is composed is extremely light, having a mean density of approximately 0.69 (water = 1) — the lowest in the Solar System. From the enormous size and extreme lightness of the globe, it has been calculated that the surface gravity



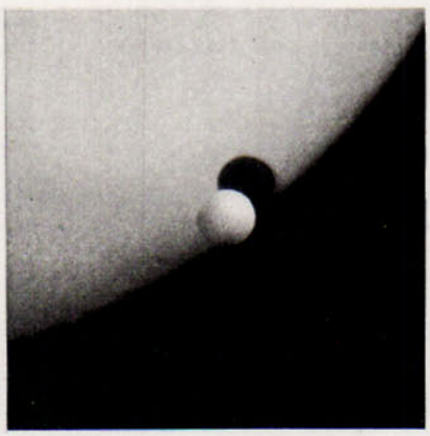


Fig. 416 — JUPITER'S SATELLITES and their shadows at the moment of opposition.

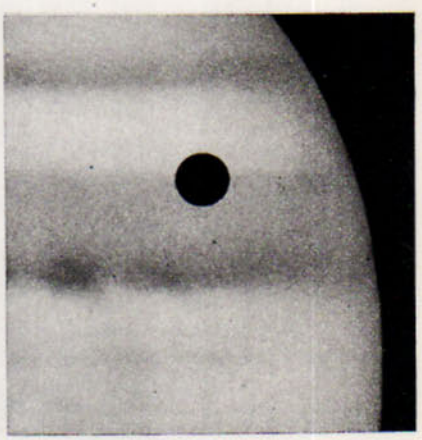


Fig. 417 — CALLISTO (IV) PASSING IN FRONT OF JUPITER

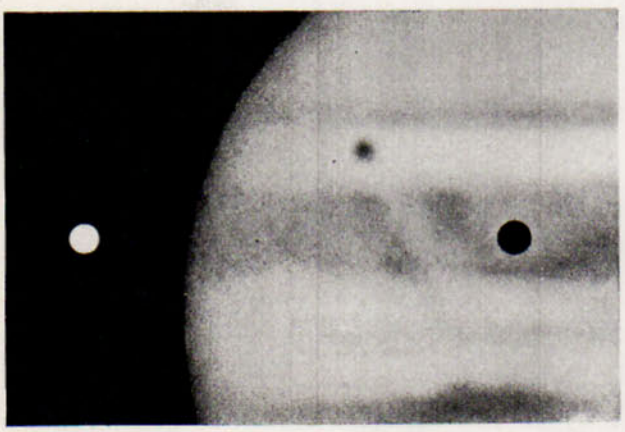


Fig. 418 — THE SHADOWS OF IO AND EUROPA ON THE DISC OF JUPITER AT QUADRATURE

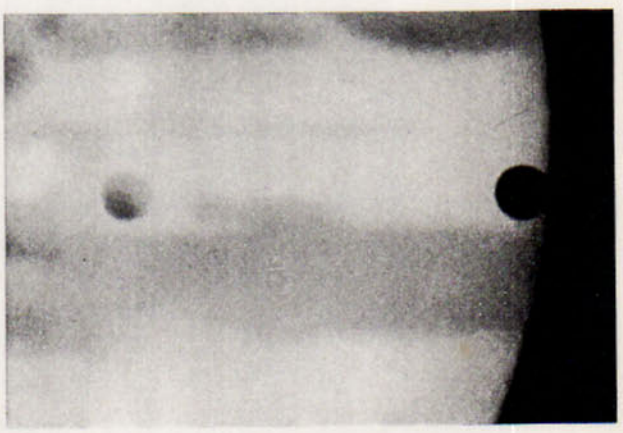


Fig. 419 — TRANSIT OF GANYMEDE AND ITS SHADOW ACROSS JUPITER

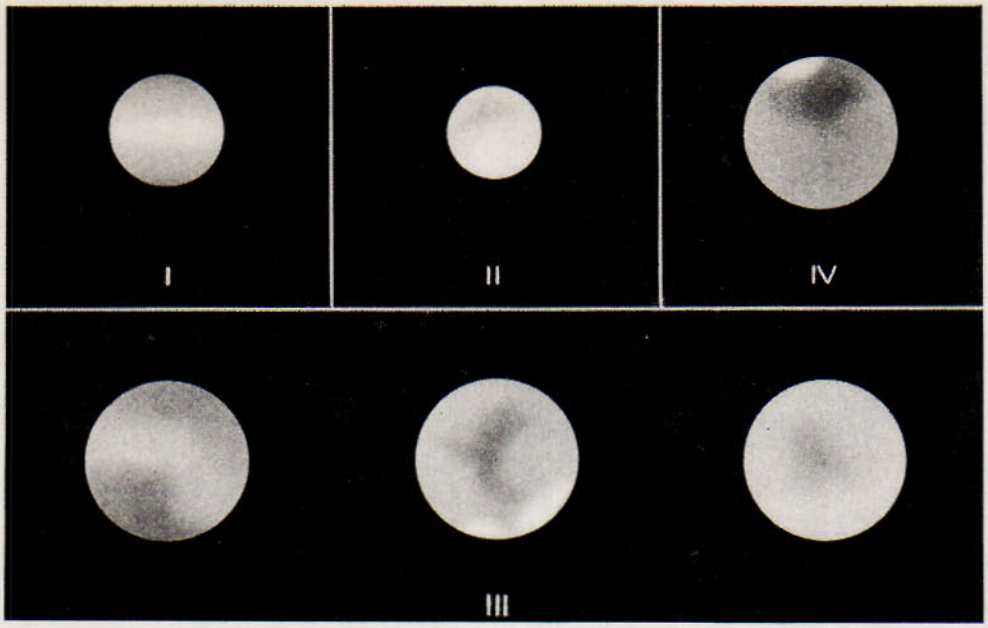
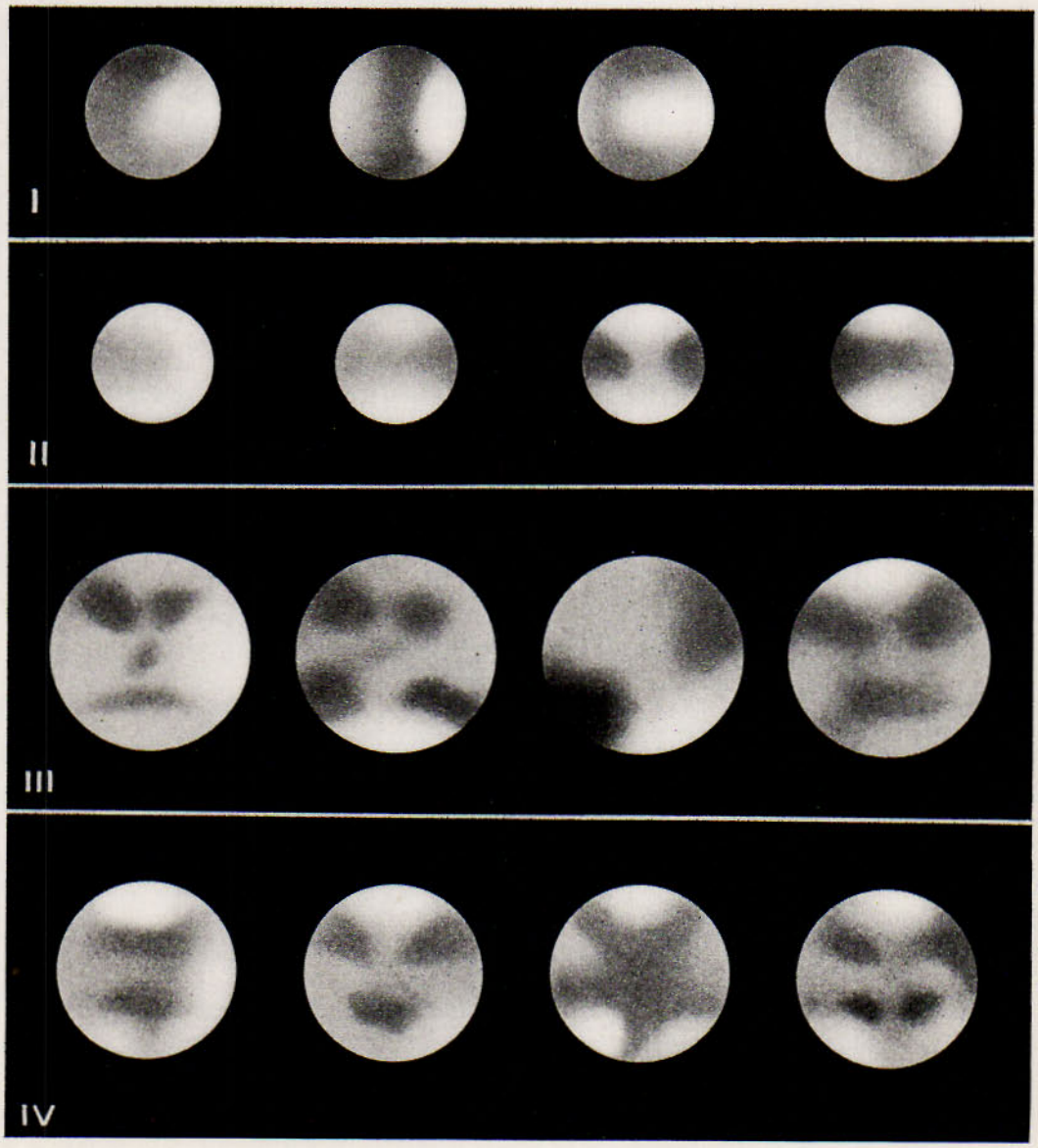


Fig. 420 — (a) JUPITER'S SATELLITES ACCORDING TO ANTONIADI



(b) AS OBSERVED AT THE PIC DU MIDI OBSERVATORY BY LYOT, CAMICHEL AND GENTILI (1941—43)

Fig. 421 — THE SIZE OF SATURN AND ITS RING SYSTEM COMPARED WITH THE EARTH

on Saturn is only just over one-tenth more than it is on Earth.

Saturn's axis of rotation is inclined at an angle of 26° 49′ to the perpendicular to the plane of its orbit. Because of its constantly changing orientation to the Sun, seasonal effects on Saturn are quite marked. Since, for these purposes, we may consider the Earth and Sun as being in the same place, the seasonal effects are revealed to us in the form of a constant change in the orientation of the planet's axis in relation to the Eearth.

Because of the long revolution period, the seasons on Saturn last about 7½ of our years, and this period also marks the changes in the orientation of the remarkable system of rings, which will be discussed in the following pages. Although the rings, which are about 170,000 miles across and situated in the plane of Saturn's equator, are independent of the planet, we cannot dissociate them, for together they form a unique system, full of interest from many points of view.

Orbit and motion of Saturn. — As already shown, Saturn revolves round the Sun at a mean distance of 886 million miles, but because its orbit is eccentric, Saturn's distance varies, between aphelion and perihelion, by just over 95 million miles, which is approximately equal to the radius of the Earth's orbit.

Saturn's orbit is so large that, even with a mean velocity of 6.03 m.p.s., the planet takes no less than 29y 168d to make one complete revolution. Consequently, from the Earth, its drift past the background of the stars is very slow, and it can be seen in much the same part of the sky from one year to the next. Saturn reaches opposition at intervals of 1 year and 12 or 13 days.

By virtue of its slow movement, Saturn returns to a given point in its orbit only after a very long interval. Its position determines the aspect under which we observe it from Earth. In the following sections, this total revolution period of about 30 years should be remembered, since it is only after such intervals that the planet repeats any particular appearance.

The appearance of Saturn. — Because of the respective sizes of the two orbits, we are never closer to Saturn than about 750 million miles, and so, in spite of its tremendous size, Saturn never appears to us as anything but a very small disc. Even at its largest this disc is only 18″ in apparent diameter, whilst the ring is only 43″ across. Thus it requires some fairly powerful optical aid in order to make out the finer features of the Saturnian system.

The way in which Saturn is tilted in relation to the Earth depends on the orientation of its axis, and therefore on its position in its orbit. Fig. 423 shows that this behaviour is precisely similar in principle to our seasons.

When Saturn is at one of its solstices, either the northern or the southern hemisphere is inclined towards the Sun and therefore towards the Earth. The same applies for the corresponding face of the ring. The latter, however, takes on an elliptical shape when tilted towards the Earth — this is simply an effect of perspective — and when tilted to its maximum, the minor axis of the ellipse is slightly larger than the polar diameter of the disc. It may be noted, incidentally, that the flattening of the globe is only noticeable when the polar axis is 'upright'. When the globe is tilted and therefore seen obliquely, its oblate appearance diminishes considerably and it looks more nearly circular.

As Saturn approaches the equinoxes, the angle at which the ring is tilted diminishes rapidly and it appears to

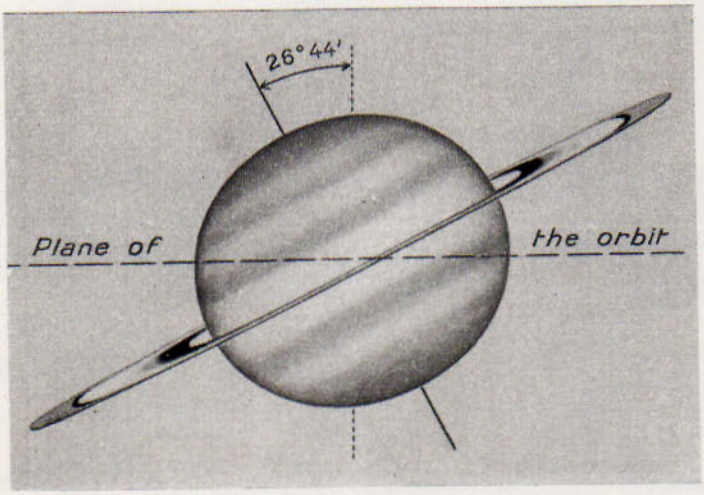


Fig. 422 — INCLINATION OF SATURN'S AXIS OF ROTATION

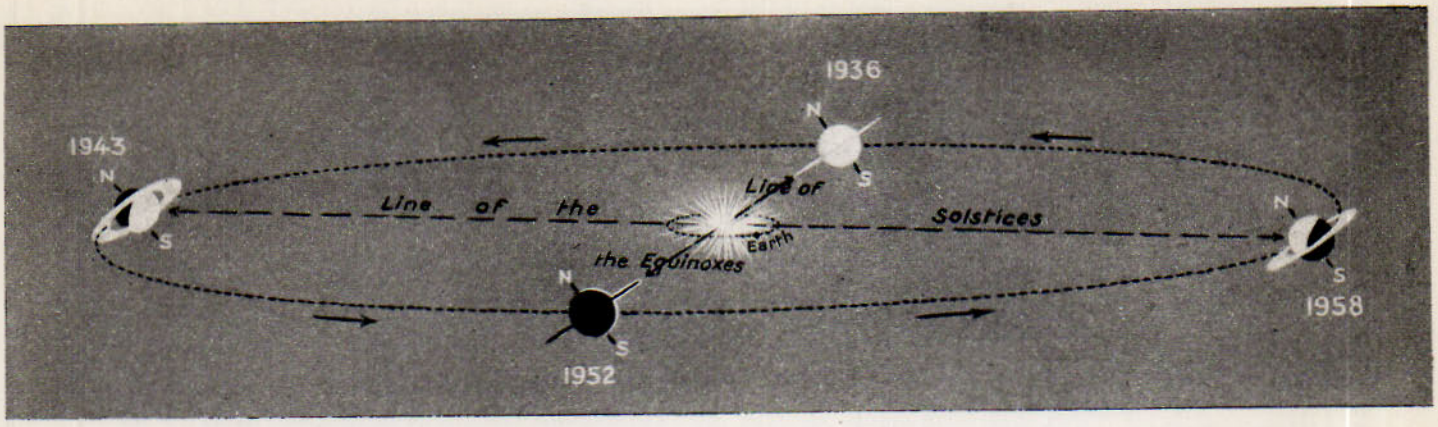


Fig. 423 — THE SEASONS ON SATURN, AND ITS POSITION IN RELATION TO THE SUN AND EARTH. We see first the north and then the south face of the ring, and between these two extremes it can be seen edgewise on at the equinoxes.

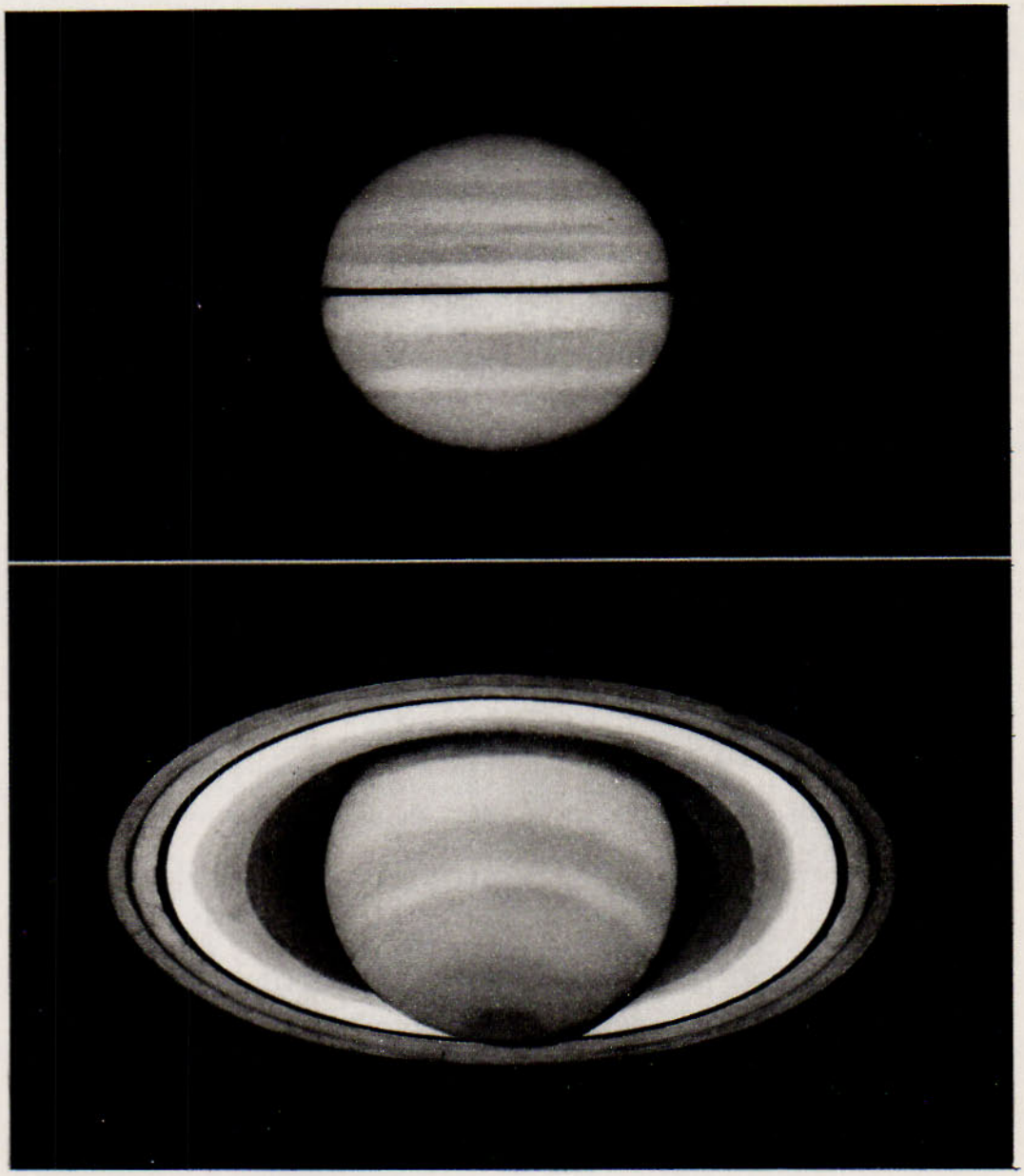


Fig. 424 — DIFFERENT ASPECTS OF SATURN AND ITS RINGS. Above: looking almost in the plane of the ring, which is nearly invisible (1936). Below: the ring is 'open' to its fullest extent (1943).

Fig. 425 — SATURN'S RINGS SEEN EDGEWISE ON, WITH TWO SATELLITES IN THE SAME PLANE

close up. At the equinoxes themselves, the plane of the ring points directly towards the Sun and Earth, so that the ring becomes nothing more than a thin line. Being a very thin formation in any case, the ring is so slender when presented edgewise on to us that it is almost invisible without powerful optical aid. It is not surprising that Saturn's habit of losing its ring threw the early astronomers into confusion.

In the above description, we have simplified the picture to some extent for the sake of clarity by considering the Sun and Earth as being in the same place. This is partly justified by their proximity, but if we wish to be strictly accurate we must allow for the Earth's distance from the Sun, as well as for the fact that the orbital planes of the Earth and Saturn do not exactly coincide. This means that the Sun and Earth will not be at the same angle to the plane of the rings.

Saturn's motion and axial inclination are measured relative to the Sun. The Earth, however, is moving relative to them both, and our movement from one side of the Sun to the other means that for us the rings disappear two or three times, whereas they disappear only once for the Sun. Fig. 426 shows why this should be so. For a little before and after the equinox, the plane of the ring no longer passes through the Sun, but through one of the positions occupied by the Earth in its orbital movement round the Sun. Consequently, as the diagram shows, the ring disappears either because we pass through its plane or because this plane is passing exactly through the Sun. In the first case, the ring is seen edgewise on, and its apparent thickness is too small to be resolved; in the second, it is still invisible, despite being slightly tilted in our line of sight, because it is illuminated only along its very thin edge. The difference between these two situations, which give much the same appearance seen from the Earth, is brought out by the very fine line of shadow that the slightly oblique illumination casts on the surface of the globe in the first case and by the absence of this shadow in the second.

These disappearances, interspersed with short periods of visibility, occur at rather long intervals. They are happenings of great interest since, among other things, they indicate that there may be variations in the thickness of the ring.

Leaving aside the smaller variations caused by the Earth's motion relative to Saturn, the above appearances are broadly repeated with a period equal to that of the seasonal changes. Between the solstices, when the face of the rings is visible, and the equinox, when the ring disappears, there is a period of just over seven years. Between this equinox and the time when the other face of the ring is visible completely, there are seven more years. Every fourteen years and a half, therefore, the ring disappears, and half way between two disappearances, it is open to its fullest in our line of sight. Every intermediate stage occurs between these two extremes, giving all the well-known features of the system, including, for example, the shadow of the globe on the ring, and of the ring on the globe.

The appearance and constitution of Saturn. — Being smaller and less brilliant than planets nearer the Sun, Saturn must be observed under good observation conditions and with reasonable telescopic power. Like Jupiter, Saturn has a series of belts running parallel to its equator, but either because of the smaller amount of light, or because of their real faintness, they do not stand out so clearly. They are also more regular in appearance than those of Jupiter. The dark and light mottlings and spots, too, are less obvious on Saturn.

At first sight, the globe, with its limb darkening, appears to be of a greenish-yellow tinge, but closer examination reveals a greater variety of shades. The bright equatorial zone is sometimes a salmon-pink colour, while the

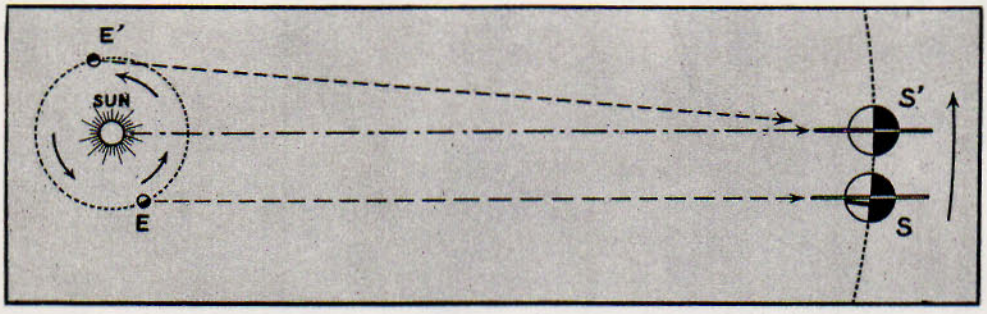


Fig. 426 — EXPLANATION OF THE DOUBLE DISAPPEARANCE OF THE RINGS when they are situated in the same plane as the Earth and Sun.

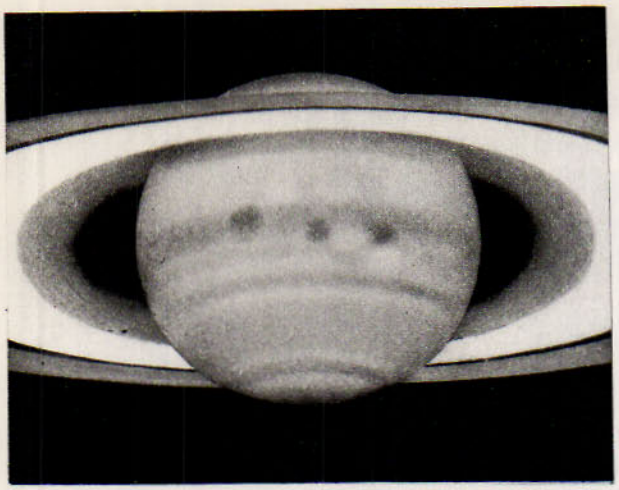


Fig. 427 — SATURN'S DISC, WITH ITS BANDS AND SPOTS

belts appear to be of a brownish-grey. The polar regions are perceptibly bluish, especially the little dark cap around the pole itself.

The features visible on the surface of Saturn are constantly varying in the same way as those of Jupiter, which may be taken as evidence for some sort of similarity in their nature. This similarity is emphasised by observations of the spots which sometimes appear on Saturn: these have shown that the speed of rotation of the surface decreases from the equator to the poles. The equatorial rotation period is 10h 12m 53s, while between latitudes 17° and 37° it is 10h 14m 45s. Other observations have yielded 10h 38m for a brilliant spot at latitude 36°. This indeed proves how unstable the surface of Saturn must be, and, as in the case of Jupiter, indicates that we are probably dealing with the outer layers of a mass of gas turning into a sort of fluid lower down.

This outer atmosphere has been investigated by spectroscopic and other methods, with interesting results.

The temperature has been estimated at approximately — 150° C, and, as on Jupiter, the spectroscope reveals the presence of ammonia and methane. Because of the extremely low temperature, it is thought that the ammonia must be in solid form for the most part, and that certain large white spots, which sometimes appear in the equatorial regions, might in fact be 'ammonia snowstorms'.

As in the case of Jupiter, it is impossible to conceive what the outer layers of Saturn must be like, since conditions are so utterly unlike anything experienced here on Earth. In any event, however, we should resist the temptation to think of Saturn as bearing any resemblance to the world upon which we live.

The appearance and constitution of Saturn's ring system. — Even a small telescope will show that Saturn possesses not just one ring, but a number of rings, each of different brightness, concentric with one another. These rings are separated by gaps, the most famous and prominent of which is known as 'Cassini's Division'. For this reason, therefore, it is more correct to speak of the *rings* of Saturn.

Broadly speaking, there are three main rings: a greyish one, known as ring A; a very bright one, known as ring B; and a dark inner ring, C, known also as the Crêpe ring, through which the globe of the planet can actually be seen. Each of these main rings is in its turn divided into smaller ones, some of which are more prominent than others, either because they are of different brightness or because they are separated by gaps. The mean brightness of the whole system is about the same as that of Saturn itself, although the ring has a whitish or silvery appearance which contrasts with the yellowish tinge of the globe.

As mentioned earlier, the total diameter of this splendid system is about 170,000 miles. Individually, the three main rings — A, B, and C — are respectively 11,000, 18,000 and 11,000 miles wide. Rings A and B are separated by Cassini's division, mentioned above, which is about 1,800 miles wide. Other gaps, less prominent, seem to exist along the dividing lines between zones of unequal brightness: the main one of these is Encke's division, which appears to separate two unequally bright parts of the outer ring.

Occasionally there seem to be irregularities in this system of rings, due to regions of unequal brightness. They are most frequently seen towards the extremities of the ellipse formed by the ring when presented obliquely, since the laws of perspective have the effect of bringing together a large number of details — such as light and dark spots — normally quite widely spaced along our line of sight.

This leads us on to study the nature of the rings. In 1659, Christian Huygens, the famous Dutch astronomer, described them in the following terms: 'a flat ring, which nowhere touches the body of the planet and is inclined to the ecliptic'. In the telescope, it does look quite solid, but the science of celestial mechanics shows that this cannot be the case: a solid ring would rapidly be destroyed by the opposing gravitational attractions of the planet itself and its retinue of satellites, attractions which are constantly varying in magnitude and direction. Nor can the ring be made of some fluid or gaseous substance, since this is not compatible with its observed opacity.

The only explanation possible is that the ring is made up of millions and millions of tiny particles, each behaving like an independent satellite. These particles are too small and too close together to be seen individually. Being independent they are free to move under the attraction of any perturbing body, whereas a solid ring would simply be broken into fragments.

This idea, which was suggested solely by the requirements of celestial mechanics, has received observational confirmation. The spectroscope allows us to measure the velocity of all the different zones across the ring system and it has been shown that this velocity decreases from the innermost out, i.e. the particles in ring C are travelling fastest, and those in ring A slowest. The inner edge of the system revolves in 4 hours and the outer edge in

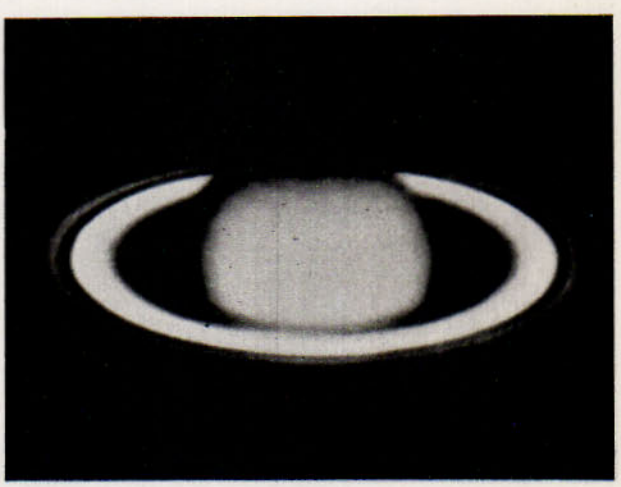


Fig. 428 — SATURN AT 0h 00m ON 20 OCTOBER 1941. Composite image from photographs taken by H. Camichel through the 15-inch telescope at the Pic du Midi Observatory.

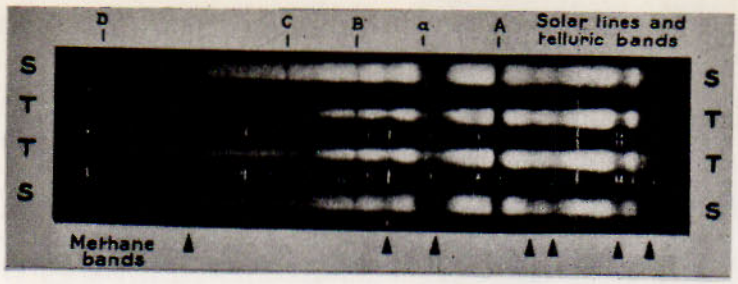


Fig. 429 — SPECTRA OF TITAN AND SATURN (RED END), showing the characteristic absorption bands of methane. *(G. P. Kuiper, MacDonald Observatory, 1944)*

14 hours. Moreover this motion can be detected by observing any irregularities in the rings, and over 150 years ago, Herschel arrived at a mean rotation period of 10h 32m.

What remains to be discovered is the actual size of the particles: are they comparable with the smaller asteroids, for example, or are they more on a par with meteoric dust? They behave, at least, like a dust formation, reflecting the light of the Sun according to how densely they are packed together. In this connection, it is not unreasonable to suppose that differences either in the actual size or the nature of the particles, or both, might make themselves felt, and these would be emphasised by any variation in the density of packing. This is suggested by the unequal transparency of different zones of the ring system: C is the most transparent, A is slightly so, whilst B is practically opaque. As a whole, the rings can be considered opaque, for they cast a completely black shadow on the planet's globe.

The theory that the rings are composed of small particles explains all the observed features, including the gaps and the differences in brightness. Moreover, the conflicting attractions of the planet and satellites ought to give rise to irregular groupings, which would explain the irregularities observed when the ring is seen edgewise on, at the time of its temporary disappearance. In this connection, it may be added that the thickness of the rings was once thought to be quite considerable, but successive estimates have continually reduced it during the last century and a half. They are now thought not to be more than 10 miles thick.

SATELLITES OF SATURN

NO.	NAME	DIAM.	DIST. IN 10^3 MLS.	PERIOD	INC.	DISCOVERY
I.	Mimas	300	113	22h 37m	26°45′	Herschel..... 1798
II.	Enceladus	400	149	1d 8h 53m	26°45′	„ 1798
III.	Tethys	600	183	1d 21h 18m	26°45′	Cassini 1684
IV.	Dione	600	235	2d 17h 41m	26°45′	„ 1684
V.	Rhea	850	328	4d 12h 25m	26°42′	„ 1672
VI.	Titan	3000	760	15d 22h 41m	26° 7′	Huyghens 1655
VII.	Hyperion	250	920	21d 6h 38m	26°	Bond..... 1648
VIII.	Iapetus	750	2,200	79d 7h 56m	16°18′	Cassini 1671
IX.	Phoebe.....	150	8,050	*550d 10h 50m	174° 7′	Pickering 1898

* denotes retrograde motion

The satellites of Saturn. — Sometimes resembling a visible continuation of the ring system, the satellites of Saturn complete the picture of this most extraordinary planet. At present, there are known to be nine, details of which will be found in the accompanying table.

It will be noted that the distribution of the satellites

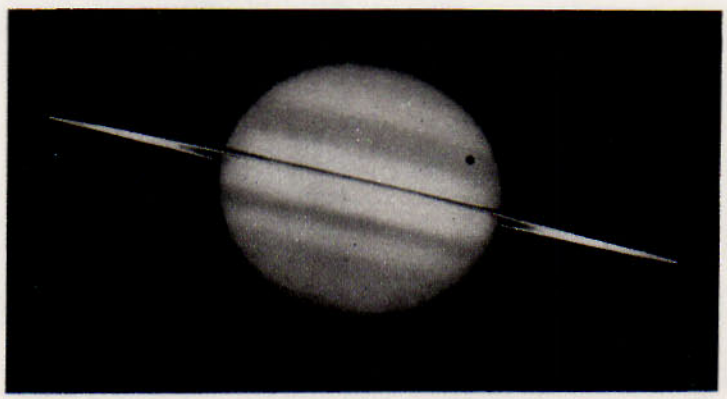


Fig. 430 — TRANSIT OF TITAN'S SHADOW ACROSS SATURN

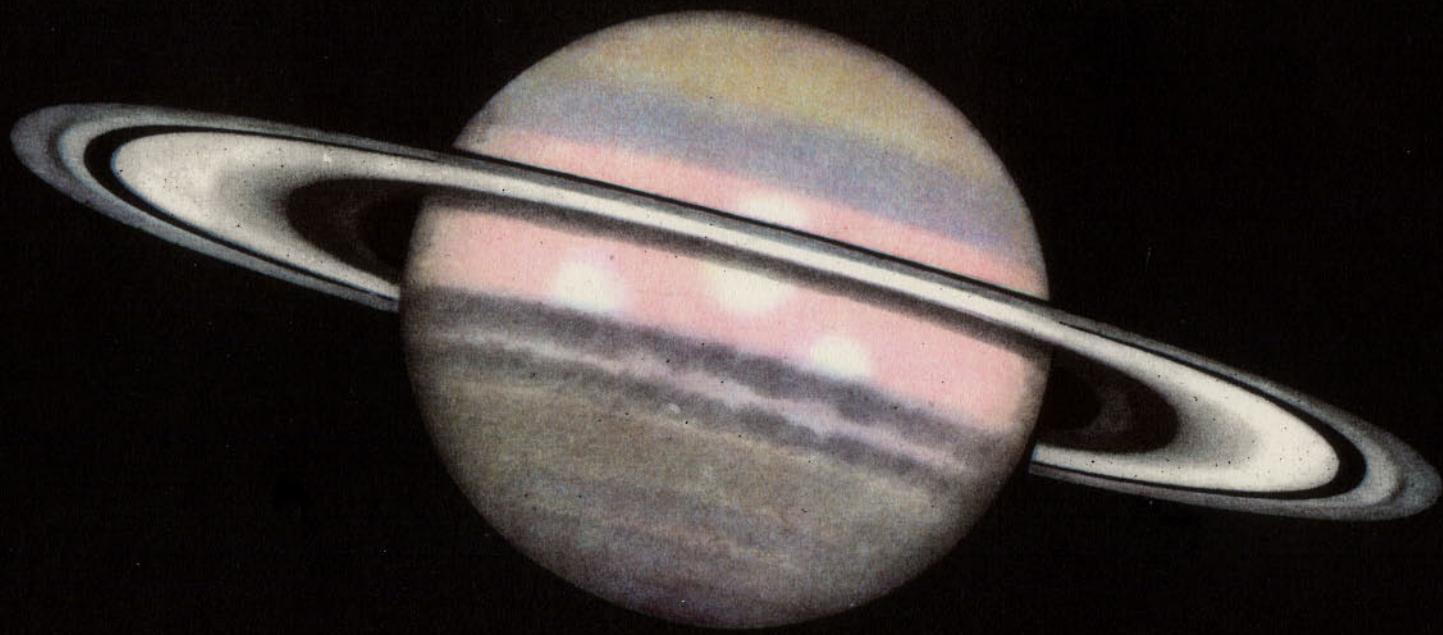


Plate V. THE PLANET SATURN

Fig. 431 — SATURN AND ITS SATELLITES IN THE FIELD OF A SMALL TELESCOPE

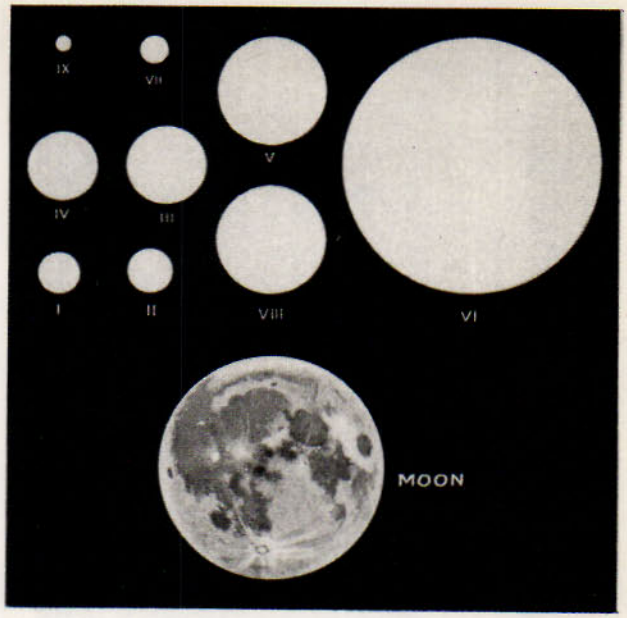


Fig. 432 — SIZES OF THE PRINCIPAL SATELLITES OF SATURN COMPARED WITH THE MOON

and the orientation of their orbits seems to divide them into two groups. The first five, relatively close to Saturn, revolve in the same plane as the rings; the other four, which are much further out, revolve in orbits which are increasingly inclined to the plane of the rings. Moreover, the orbits of VII (Hyperion) and IX (Phoebe) are highly eccentric, the latter moving in the retrograde direction.

With the exception of Titan, which can be seen in a fairly small telescope, and Iapetus and Rhea, which are also fairly easy to see, it requires considerable optical power to pick out these bodies. As in the case of Jupiter, the last of Saturn's satellites was discovered only by photography. For the most part, their dimensions are known from photometric studies, but only approximately. Titan is the only really large one.

Even if the satellites were always observable, they would rarely be seen to transit the planet, casting a shadow on the globe, or be eclipsed themselves in the planet's shadow. This is because the laws of perspective work in the same way for the satellites as for the rings. The satellites are situated almost in the plane of the rings; when the latter are tilted, the orbits of the satellites assume a similar shape and the satellites move along paths which are apparently parallel with the edge of the ring system. This means that under such conditions they will always pass either above or below the disc, but never in front of it. It is only at the rare intervals when Saturn is at the equinoxes that the orbits of the satellites, like the rings, are presented edgewise on to us and the Saturnian moons pass in front of the planet, casting a shadow on the disc.

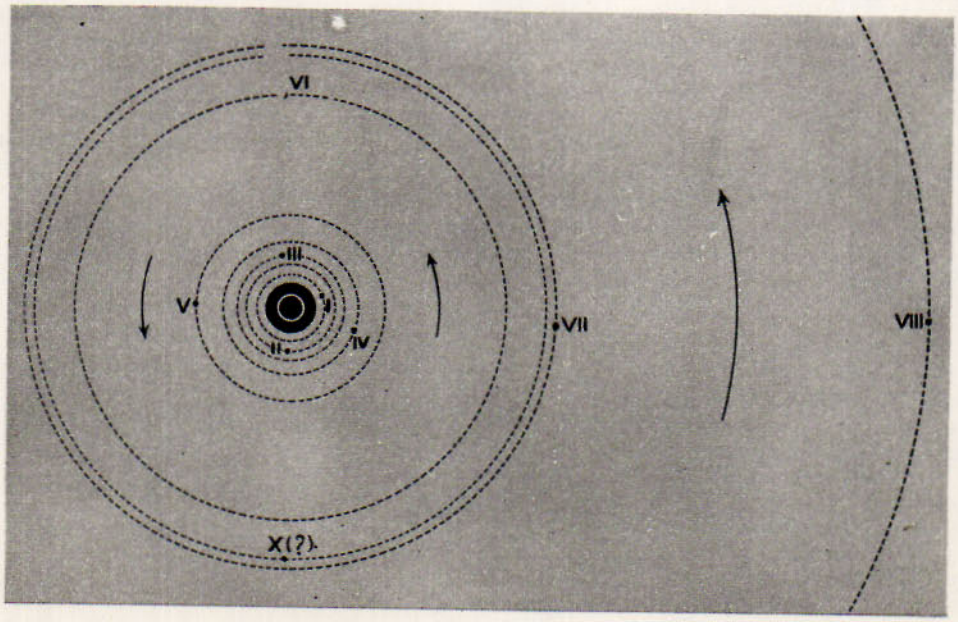


Fig. 433 — ORBITS OF THE PRINCIPAL SATELLITES OF SATURN

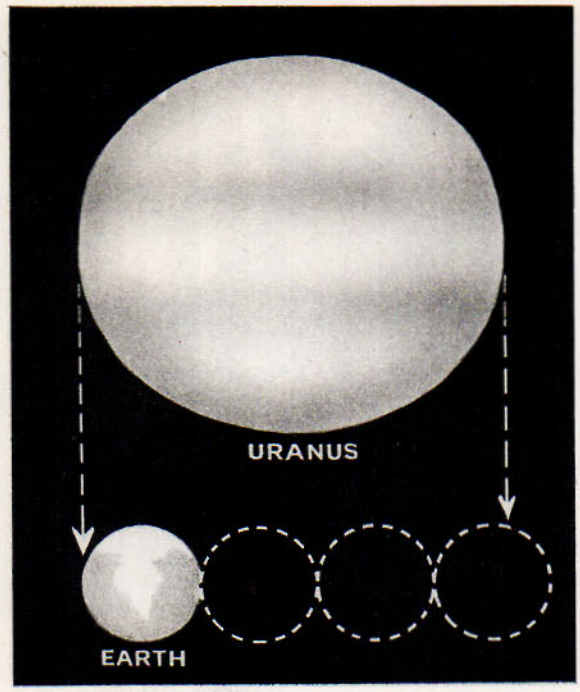


Fig. 434 — THE SIZE OF URANUS COMPARED WITH THE EARTH

This occurs only every 15 years. Titan, moreover, is the only one of Saturn's moons large enough to cast a shadow visible from the Earth.

The sole indication as to the physical nature of these bodies is furnished by the slight changes in brightness that they apparently undergo at different points on their orbits. For example, the regularity of the fluctuations in the brightness of Titan and Iapetus shows that they must make one complete rotation in the time that they take to revolve round Saturn. This is what happens in the case of the Moon and the satellites of Jupiter, which makes it probable that this condition is a normal one among satellites generally.

Some remarkable spectroscopic work, carried out in the U. S. A. by Kuiper, has shown that Titan possesses an atmosphere closely resembling that of Saturn itself, i.e. composed principally of methane and ammonia. This is the first recorded case of a satellite having an atmosphere of any importance. It may be noted that the same very sensitive methods applied to the system of Jupiter have revealed nothing comparable; this proves that none of Jupiter's satellites can have an atmosphere of any great density. The similarity between the atmospheres of Saturn and Titan would indicate some close relation between their modes of origin, and this fact is very significant in considering the origin of the Solar System itself.

URANUS

Dimensions of the globe. — Although classed among the giants of the Solar System, Uranus in considerably smaller than either Jupiter or Saturn. Its diameter has been measured as 29,300 miles, which is under 4 times that of the Earth. The volumes of the two globes are therefore in the ratio of about 60 to 1.

Like Jupiter and Saturn, Uranus has a very low density, which, from its mass and volume, can be calculated as 1.27 (water = 1). The similarity between the giant planets is strengthened by the fact that the globe of Uranus is flattened at the poles; the oblateness has not been determined with complete accuracy but appears to be about 1/12 or 1/14. This is due to the very rapid rotation, which appears to have a period of about 10h 41m. The axis of rotation, however, is unique, since it is inclined at an angle of 98° to the perpendicular to the plane of the Solar System — in other words, whereas the axis of the Earth is tilted only 23° from the 'vertical', Uranus's axis has tilted by *more than a right angle*, so that one might say the planet goes round its orbit 'lying on its side'.

It will be noted that many of the above values are given as approximations. This uncertainty is due to the great distance of the planet, which means that at opposition Uranus is no more than a tiny disc of magnitude 5.7, so that exact measurement is a matter of some difficulty.

Orbit and motion of Uranus. — Uranus revolves at a mean distance of 1,783 million miles, but because of the eccentricity of its orbit, its distance from the Sun may vary by 166 million miles between aphelion and perihelion.

Being so far away from its primary, Uranus moves extremely slowly. Its mean velocity is 4.25 m.p.s., which means that the planet accomplishes one complete revolution of its immense orbit in 84y 7d. Thus a terrestrial astronomer who observes Uranus at a given position in its orbit is unlikely to be still alive at its next return to that point.

Each of Uranus's seasons lasts for 21 ½ of our years, on average. They are very pronounced because of the high inclination of the planet's axis.

Uranus travels along an orbit which is not very steeply inclined to the plane of the ecliptic, the angle between them being only 0° 46′. With its slow movement, therefore, it moves round the sky along much the same path as the Sun, and comes to opposition every 369.7 days.

Appearance and constitution of Uranus. — Being so far away, Uranus is practically invisible to the naked eye. Nevertheless, when favourably placed in a very clear sky, it can be seen by a keen-sighted person, since it is about equal in brightness to the faintest star which can normally be seen under such conditions.

At such a great distance, the illumination that Uranus receives from the Sun is comparatively feeble, and, as already mentioned, the planet presents only a very small disc, never exceeding 4″ at the most.

This small size, coupled with the planet's comparative faintness, requires the use of powerful instruments if the image is to be studied to good effect. Under satisfactory conditions, the disc of Uranus appears to be of a greenish

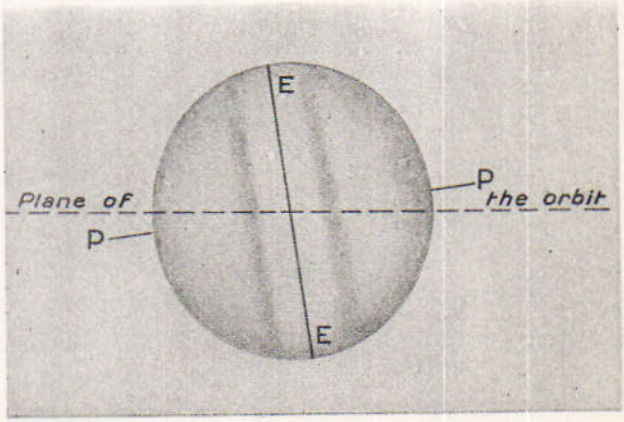


Fig. 435 — INCLINATION OF URANUS'S AXIS OF ROTATION

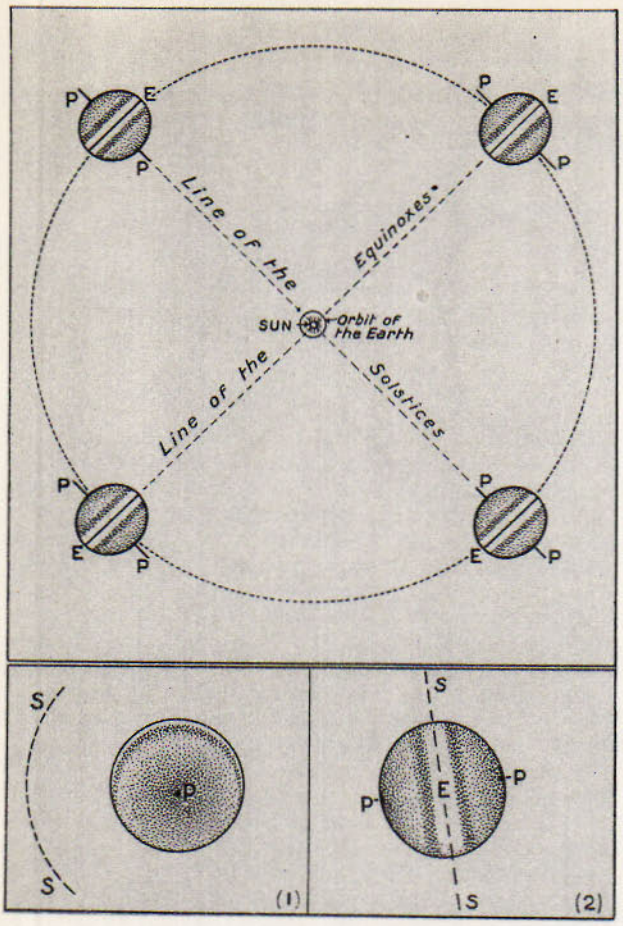


Fig. 436 — THE SEASONS ON URANUS, and the changing orientation of its globe with respect to the Earth.

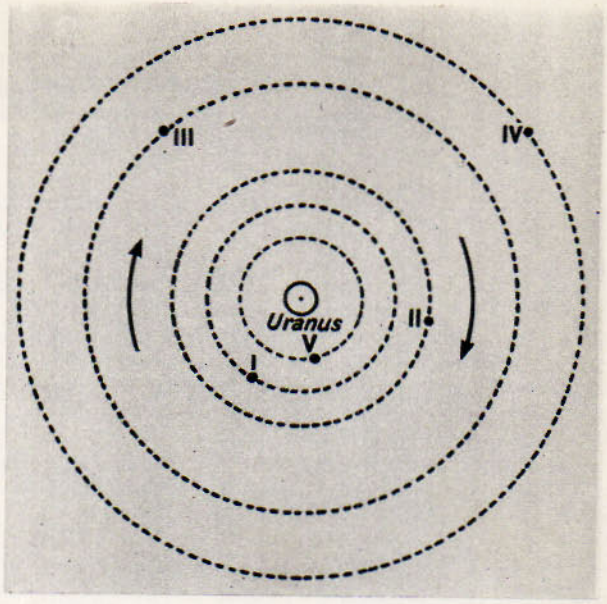


Fig. 437 — ORBITS OF URANUS'S SATELLITES

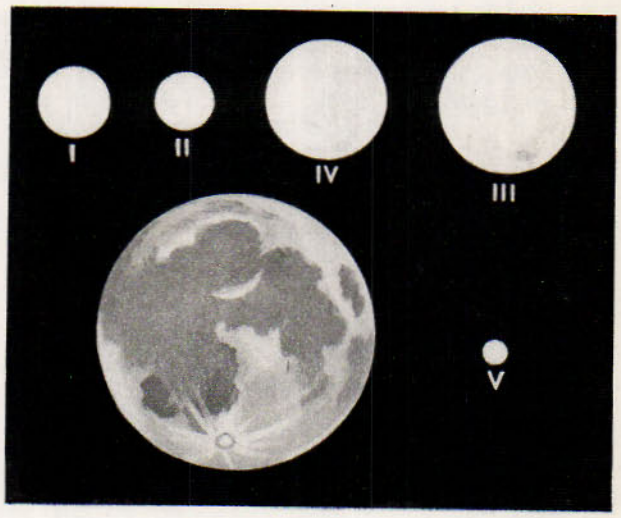


Fig. 438 — THE SIZES OF URANUS'S SATELLITES COMPARED WITH THE MOON

blue colour; like those of Jupiter and Saturn, it is slightly darker round the limb, and has a system of faint dark belts parallel with the equator. These belts are greyish in colour, and do not show any structural details. The apparent simplicity of the picture is, however, probably due to the faintness of the telescopic image rather than to any real paucity of detail. At any rate, the little that we can see indicates that Uranus — as can be deduced from the density, shape and rapid rotation of the globe — is essentially similar in structure to Jupiter and Saturn.

The difficulties of observing Uranus stem not only from the general faintness of the planet but also from the peculiar positions in which it is presented to us. The inclination of the planet's axis — lying almost within the plane of the Solar System — means that at the solstices one or other of Uranus's poles is pointing directly towards us, and therefore appears almost in the centre of the disc. The disc itself will look circular because it is the equator, not the meridian, which is facing us. The belts are almost invisible, being compressed by perspective around the edge of the disc.

These considerable differences are somewhat difficult to appreciate properly because of the delicate nature of

SATELLITES OF URANUS

NO.	NAME	DIAM.	DIST. IN 10^3 MLS.	PERIOD	INC.	DISCOVERY
V.	Miranda	100?	76	1d 9h 50m	97°59′	Kuiper 1948
I.	Ariel.....	1500	119	2d 12h 29m	97°59′	Lassell 1851
II.	Umbriel.....	800	166	4d 3h 28m	97°59′	,, ... 1851
III.	Titania.....	1500	272	8d 16h 56m	97°59′	Herschel..... 1787
IV.	Oberon	1500	364	13d 11h 7m	97°59′	,, ... 1787

N. B. There is considerable uncertainty attached to the values for the diameters.

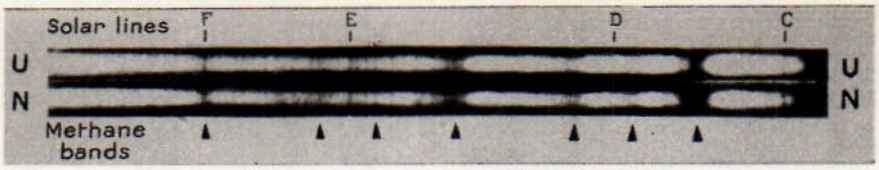


Fig. 439 — SPECTRA OF URANUS AND NEPTUNE, showing the characteristic absorption bands of methane. *(G. P. Kuiper, MacDonald Observatory, 1944)*

the image, mentioned above. They help to explain the sometimes conflicting results announced by observers from the beginning of the nineteenth century onwards.

All the similarities revealed between the physical constitution of Jupiter and Saturn, the shape, density and appearance of their globes, have been confirmed by spectroscopic study of Uranus. It appears that the atmosphere is predominently composed of methane. Ammonia, if its exists, must be in the solid state, for the temperature is about −160°C.

The satellites of Uranus. — At present there are 5 known satellites of Uranus, which are all difficult to observe, since, being so far away from the Sun, they are very faint. They can be seen only in powerful telescopes, and the last one, Miranda, was only discovered photographically.

Details of each satellite are given in the table at the foot of page 225.

All the satellites of Uranus move along almost circular orbits in the retrograde direction. This is because they revolve in the plane of Uranus's equator, which, as we have seen, is tilted by more than a right angle. Thus, when the movements of the satellites are projected on to the plane of the Solar System, they appear retrograde. Like the planet, the orbits of the satellites are sometimes presented edgewise on, sometimes face on, to the Earth.

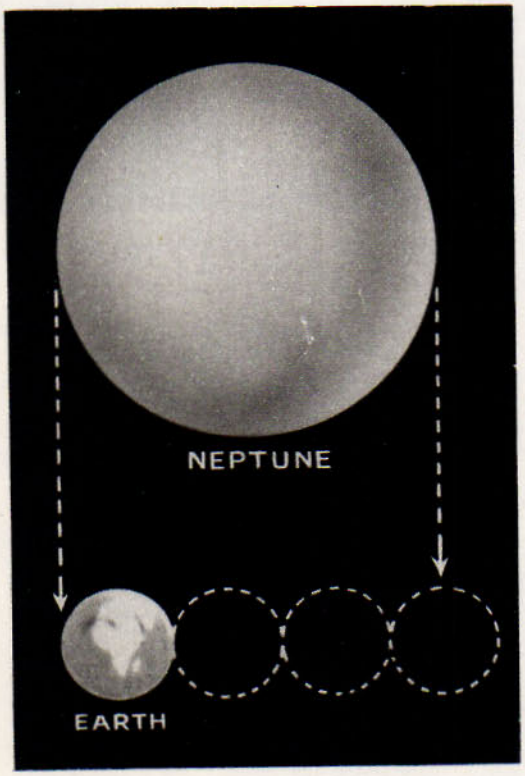


Fig. 440 — THE SIZE OF NEPTUNE COMPARED WITH THE EARTH

Nothing is at present known about the physical constitution of these satellites; the dimensions indicated in the table are only approximations.

NEPTUNE

Dimensions of the globe. — One of the family of giant planets, Neptune is so similar to Uranus in size that first one and then the other has been thought to be the larger. This confusion arises from the difficulty of making exact measurements of such tiny discs. Latest results indicate that Neptune is slightly the smaller of the two, with a diameter of 27,700 miles against Uranus's 29,300. There is, however, a considerable degree of uncertainty still attending these values, since different authorities quote different figures.*

The mass of Neptune is calculated to be 17.3 times that of the Earth; its density is therefore 0.41 of the Earth's, and its specific gravity 1.5.

Neptune rotates in a period of about 15h 40m (derived from spectroscopic measurements) about an axis inclined 29° to the perpendicular to the plane of the ecliptic. The rapid rotation must render the planet's globe appreciably oblate, which is confirmed by the precessional movement observed in the orbit of its largest satellite.

Orbit and motion of Neptune. — Neptune's immense orbit is almost circular, having an eccentricity of only 0.0087. Its inclination is 0° 47′ to the plane of the ecliptic. The planet's mean distance from the Sun is 2,793 million miles, and at a mean velocity of 3.4 m.p.s. Neptune takes 164y 280d to make one complete revolution. Its movement in the sky is consequently very slow, and it comes to opposition at intervals of only 367 days, taking years to pass through a single constellation.

For those who are not familiar with planetary movements, it may be pointed out that, because of its extremely slow pace, Neptune has not yet been observed through one complete circuit of the celestial sphere. The planet will not return to the place where it was discovered until the year 2011!

* The values used above are those given in the B. A. A. Handbook.

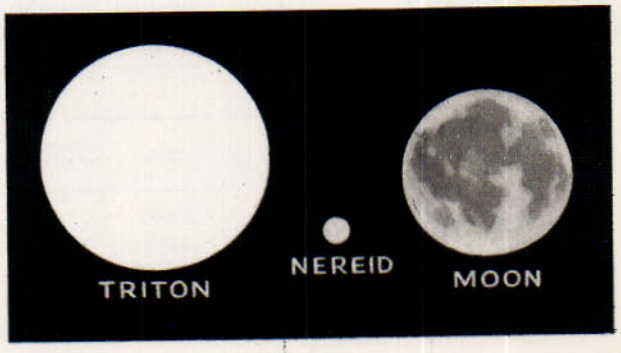


Fig. 441 — THE SIZES OF NEPTUNE'S SATELLITES COMPARED WITH THE MOON

Appearance and constitution of Neptune. — Because of its great distance from both the Sun and Earth, the annual variations in Neptune's distance from our world has no appreciable effect, and its apparent diameter may be regarded as constant at 2.4″.

This small size and the low illumination received from the Sun render Neptune a very faint object in our skies, quite invisible to the naked eye. A telescope of 4-inch aperture or more is required to reveal its true nature.

The minute disc of the planet is of a greenish-blue colour, rather similar to that of Uranus, but very little detail can be made out. There appear to be dark belts running parallel with the equator, which reinforces Neptune's similarity with the other giant planets. This similarity is further confirmed by spectroscopic work, which reveals the same atmospheric composition: the predominance of methane. It is possible that because of the low temperature of −170°C much of the methane has turned to liquid, whilst most other gases known to us must have solidified.

The satellites of Neptune. — At present, Neptune is known to have 2 satellites: the larger, Triton, revolving at a distance of 220,000 miles from its primary, was discovered by Lassell in 1846, whilst the second, Nereid, was discovered as recently as 1949, by Kuiper.

The motion of Triton is retrograde. Its orbit is inclined by 20° to the planet's equator, and it precesses in a period of about 580 years, owing to Neptune's equatorial bulge. The diameter of this satellite is about 3,300 miles, and to see it at such a tremendous distance requires a fairly powerful telescope.

Nereid, a very small, faint body only about 200 miles in diameter, revolves in an extremely eccentric orbit. Its distance from Neptune varies from 828,170 miles to 6,086,551 miles, and it takes about 359 days to make one complete revolution. Its motion, unlike that of Triton, is direct. Its low magnitude, 19.5, puts it beyond the power of any but the largest telescopes.

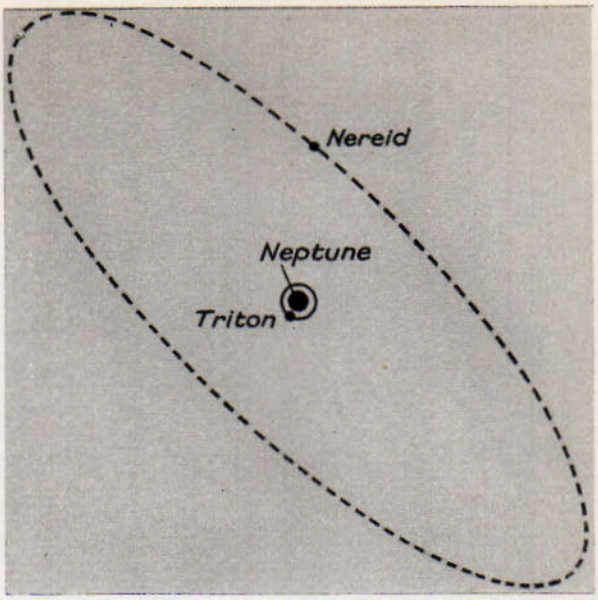


Fig. 442 — THE ORBITS OF NEPTUNE'S SATELLITES

PLUTO

We have very little reliable information about Pluto, whose orbit forms the furthest known boundary of the Solar System.

The orbit of Pluto is remarkable in a number of ways. Its eccentricity is the highest (excepting the asteroids) of any in the Solar System — 0.247. Consequently, although its mean distance is 3,666 million miles, Pluto can come within 2,766 million miles of the Sun at its nearest and is 4,566 million miles away at its farthest. Its orbit is inclined at an angle of 17° 6′ to the plane of the

SATELLITES OF NEPTUNE

NO.	NAME	DIAM.	DIST. IN 10^3 MLS.	PERIOD	INC.	DISCOVERY
I.	Triton	3,300	220	*5d 21h 3m	139°	Lassell 1846
II.	Nereid	200?	3,500	359d	48°	Kuiper 1949

* denotes retrograde motion

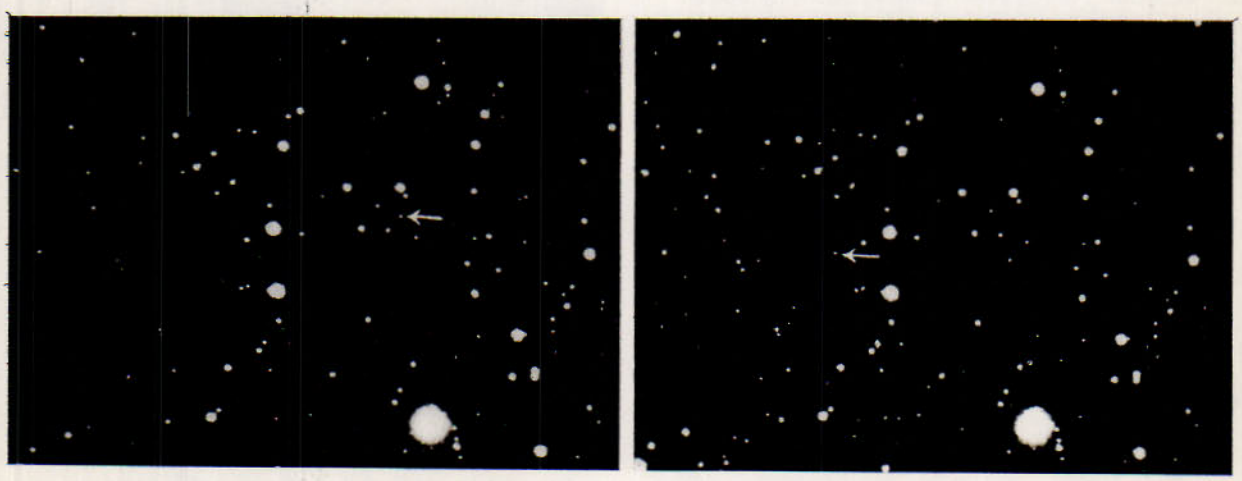


Fig. 443 — THE DISCOVERY OF PLUTO, detected from its movement between the position on the left (23 January 1930) and that on the right (29 January 1930). *(Photographs, C. W. Tombaugh, Lowell Observatory)*

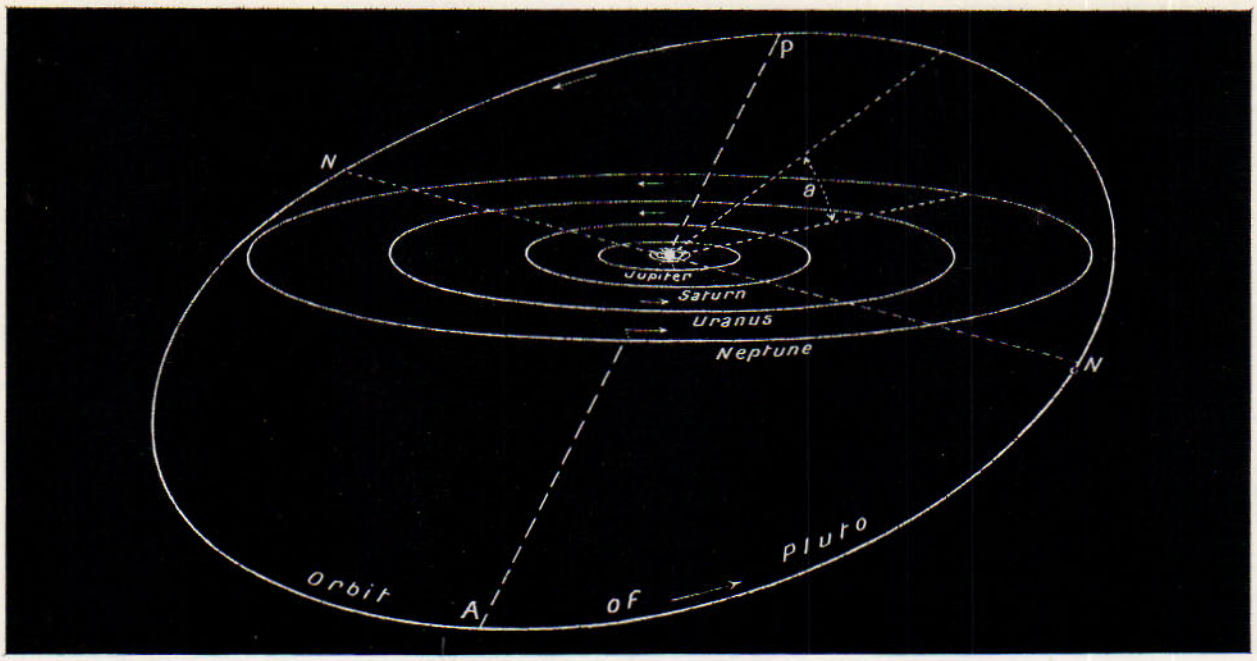


Fig. 444 — A PERSPECTIVE VIEW OF PLUTO'S ORBIT, which is steeply inclined to all others in the Solar System. N-N, line of the nodes, i. e. the intersection of the different planes, the angle they make between them being indicated *a*. (P, perihelion; A, aphelion.)

ecliptic, which is exceeded only by a few minor planets and asteroids.

It will be noted that at perihelion Pluto is a little nearer the Sun than Neptune. Since its discovery in 1930, Pluto has been travelling towards perihelion, but it will not arrive there until 1989. Consequently, astronomers at the end of the century will be in a slightly better position to study the planet than those alive today. Pluto will not return to the place where it was discovered until the year 2177, since it has a revolution period of 247.7 years.

This 'Ultima Thule' of the Solar System is such a faint object that a large telescope is required to observe it visually, although its position is usually measured on long-exposure photographs. Its diameter cannot be measured directly, but Kuiper, using a disc-meter on the 200-inch telescope, estimated that it is 0.46 that of the Earth. Pluto therefore has about one-tenth of the Earth's volume. The mass has been tentatively deduced from perturbations, and appears to be comparable with that of our own planet, which gives Pluto a density of more than 10 times the Earth's. This is difficult to accept, and it underlines the uncertainty of the values with which we are dealing.

It has been suggested that the reason for this lies in a peculiarity in the way Pluto reflects light. The apparent brightness of a planet depends on the reflecting power of its surface and its distance from the Sun and Earth. Thus a given brightness could result from a very small globe with high reflecting power, or a larger one with a low albedo.

Applying this principle to Pluto, attempts have been made to estimate its diameter taking a constant magnitude and different reflecting powers, corresponding to those of the various types of planet in the Solar System. The results give a maximum diameter of only 0.80 of the Earth's and a minimum about equal to the diameter of the Moon. The mean value corresponds to a planet about the size of Mercury.

Another suggestion is that Pluto is covered with a liquefied or frozen atmosphere, possibly smooth enough to act as a sort of mirror. We should then be measuring not the planet's diameter but the diameter of an image of the Sun. This seems unlikely in view of the planet's low magnitude. A more plausible version of this idea is that the surface is mottled, and that we are measuring the diameter of a large white patch. Only the future will resolve this problem.

Because of its small size and mass, Pluto belongs to the group of terrestrial planets — Mercury, Venus, Earth and Mars — rather than the gaseous giants. It is possible, therefore, that Pluto is not the last planet in the Solar System, but merely the first of another group yet to be discovered. It would be rash to dismiss this possibility, knowing how powerful astronomical instruments have become over the last few years, but the discovery of planets beyond Pluto would be an arduous task indeed.

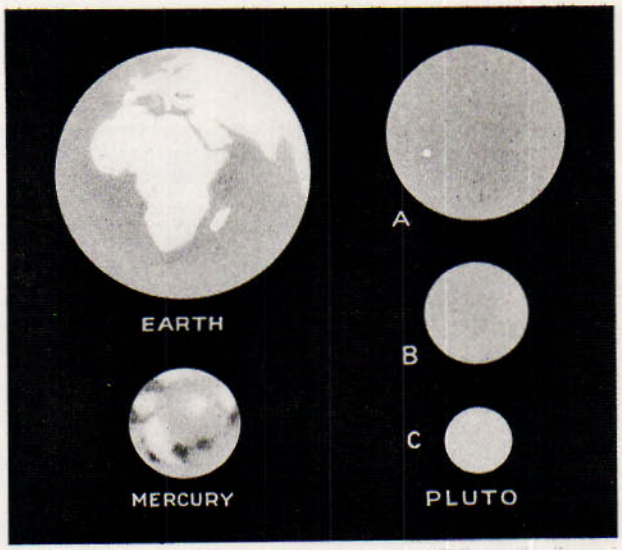


Fig. 445 — POSSIBLE SIZE OF PLUTO, compared with Mercury and the Earth. A, B and C are estimates based on various possible powers of reflection, B being the mean and A and C the extremes.

Fig. 446 — THE GREAT COMET OF 1843

CHAPTER 8

COMETS

Of all the heavenly bodies visible to the unaided eye, comets are the most remarkable. Their frequently sudden and unheralded appearances in our skies, the abnormal paths they take across the celestial sphere, and the extraordinary spectacle that a large comet can produce, make them altogether exceptional.

Although the origin and precise nature of these objects is still in some doubt, comets obey the same laws as the other members of the Solar System, moving in highly elliptical orbits, which are frequently inclined at a considerable angle to the orbits of the planets. Their movement may be either direct or retrograde. The reason that the path of a comet appears to be so capricious in relation to the planets is only because of the comet's unusual and constantly changing position relative to the Earth. Thus a comet may appear suddenly in any direction from the depths of space, rush past the Earth and round the Sun, and disappear in possibly quite another part of the sky. This occurred, for example, in 1957, which saw the appearance of two bright new comets, Arend-Roland and Mrkos. During their comparatively short period of visibility, comets sometimes change remarkably in appearance; this is due both to the effect of perspective and to real changes in the comet itself, which are caused by the Sun.

Cometary orbits and motions. — Cometary orbits differ from planetary orbits by being very elongated, and on occasion steeply inclined to the plane of the Solar System. The high eccentricity causes considerable variations in the comet's distance from the Sun.

A comet's orbit may be one of three different geometrical curves: an ellipse, a parabola or a hyperbola. An ellipse is a closed curve, whereas a parabola and a hyperbola are curves whose arms become further and further apart, without ever closing up again.

In Chapter 1, we saw that an ellipse is constructed round two foci. The eccentricity of the ellipse is the distance to the centre of the figure from one of the foci, expressed as a fraction of the semi-major axis. The parabola may be compared with an ellipse, one of whose foci is placed at infinity. Alternatively, one can say that the eccentricity of a circle is zero (because the two foci coincide with the centre), whilst in an ellipse the higher the eccentricity, the more elongated the figure. When the eccentricity becomes equal to 1, one focus is removed to infinity and the figure becomes a parabola. The hyperbola is an even more open curve, the eccentricity of which is greater than unity.

The path followed by one object about a larger one which is attracting it gravitationally is determined by the

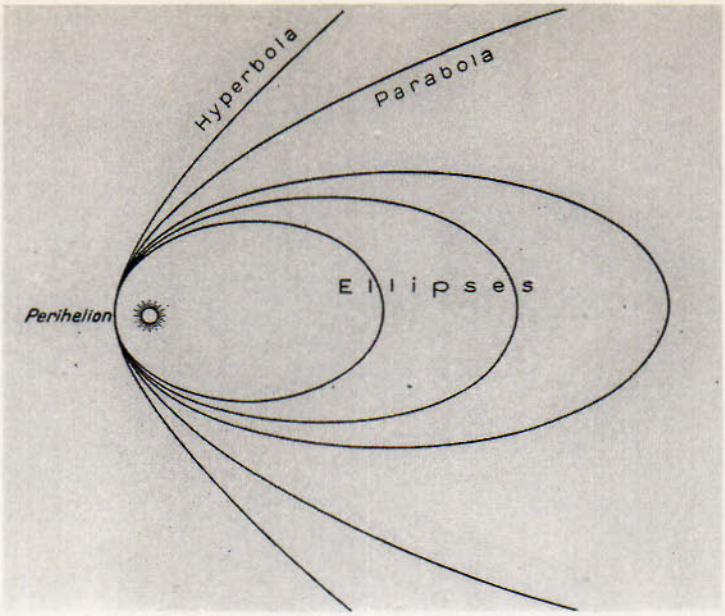


Fig. 447 — At perihelion, the different types of cometary orbit are often indistinguishable.

relative speed. At a certain speed, called 'circular velocity', the first body will move uniformly in a circle round the second. With a slightly higher speed, its path will be in the form of an ellipse, and the greater the object's velocity the higher the eccentricity of the ellipse. If this speed is to circular velocity as 1.414 is to 1, then the object's path will be parabolic; if greater, it will be hyperbolic.

Given this difference between the various curves, in the first case the comet travels along a closed path and so returns periodically, like a planet, to the same position in relation to the Sun. In the second case, the comet arrives from the depths of space, passes through the Solar System and around the Sun, and finally goes off into space in another direction. In other words, comets may be divided into two categories: periodic, which return after a given interval, and temporary visitors to the Solar System which pass through the vicinity of the Sun only once. It should be noted that the three sorts of curves are very difficult to distinguish from observation of their limited arcs, since at perihelion they are often practically identical (Fig. 447).

It is generally accepted that the majority of comet orbits are either elliptical or parabolic. Some are slightly hyperbolic, at least within the Solar System, but when it is possible to analyse their movements completely it is found that outside the Solar System they are parabolic. How can we explain this?

It should be noted that the above remarks are purely theoretical, in that they relate to a comet moving under the influence of the Sun alone. However, the influence of

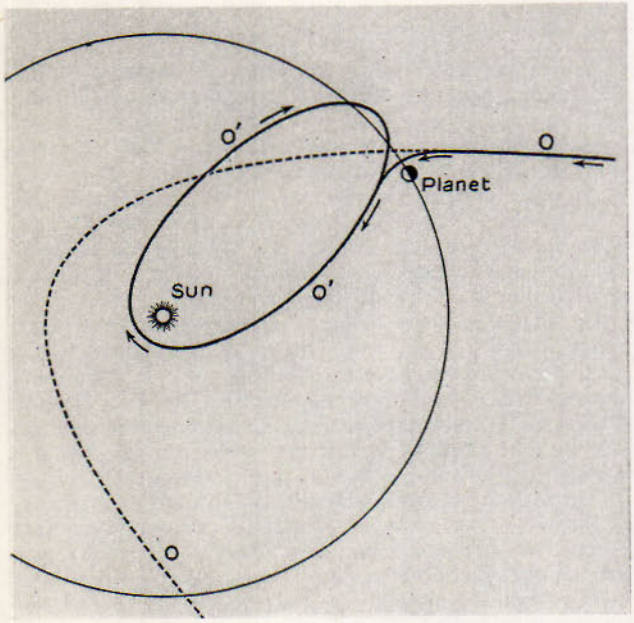


Fig. 448 — CAPTURE OF A COMET BY A MAJOR PLANET. O, original orbit; O', new orbit.

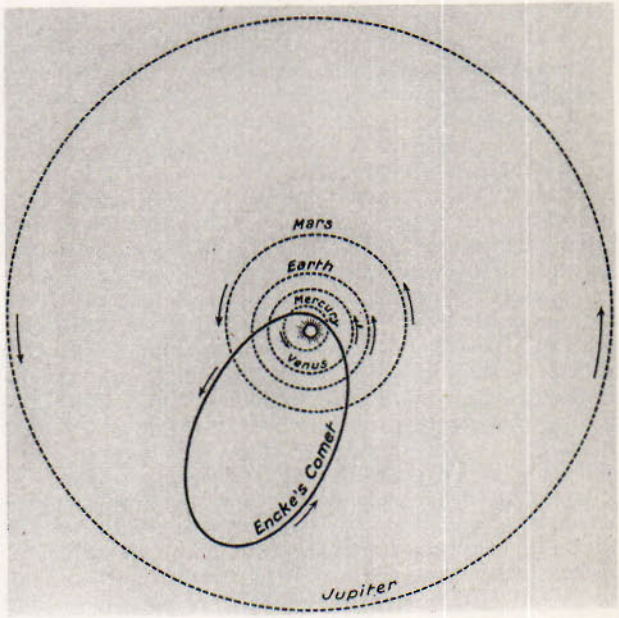


Fig. 449 — ORBIT OF ENCKE'S COMET

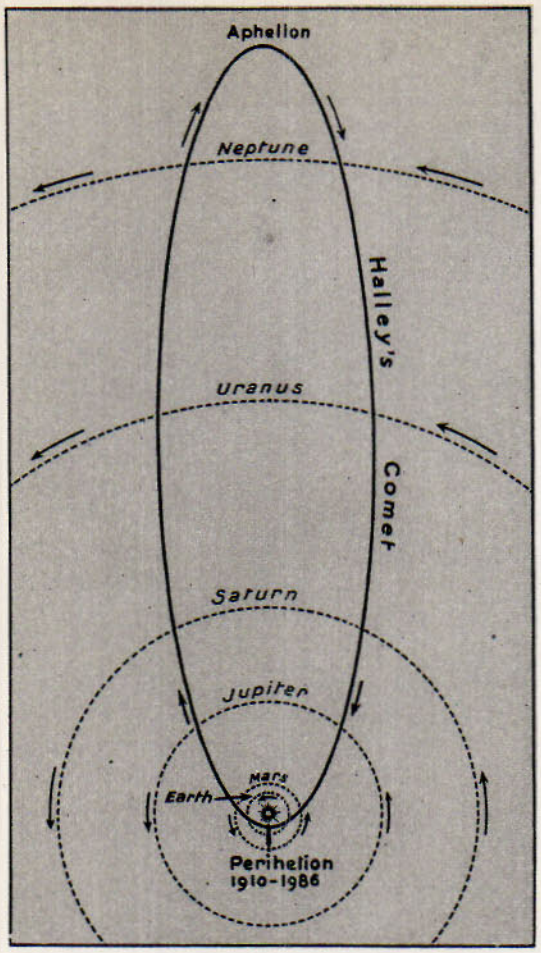


Fig. 450 — ORBIT OF HALLEY'S COMET

the planets cannot be disregarded, because, although small in comparison with the central Sun, their mass is sufficient to have effect. Moreover, the orbits of some comets are so orientated that they intersect the orbit of a planet, in which case there may ultimately be a very close approach between the two bodies. This possibility of collision may explain the terror with which people were once seized upon the appearance of a bright comet in the sky. Later in the chapter, it will be shown what probability there is of this happening. A close approach between a comet and a planet, however, may cause perturbations in the motion of the comet sufficient on occasion to alter the shape of its orbit completely.

Because of these perturbations, therefore, a comet's revolution period may be considerably lengthened or shortened. Moreover, it is known that a number of periodic comets have not always followed their present orbits, these being the result of 'captures' at some previous passage across the Solar System. Fig. 448 gives an example of how this might occur. The comet is originally travelling along a parabolic path, OO, but it passes near a planet, whose gravitational attraction is sufficient to pull it off course into the new elliptical orbit, O′ O′. The aphelion point of this orbit, it may be noted, is situated near the place where the capture occurred. Of course, this could also occur the other way round, with a periodic comet being pulled out of its ellipse into a parabolic or hyperbolic orbit, in which case it would leave the Solar System altogether.

It has been suggested that all periodic comets at present known to be in the Solar System were captured in this way by one of the major planets — Jupiter in particular. This, however, is a question which has not yet been settled.

The behaviour of comets is interesting. As a rule, comets become observable only when they are less than twice the Earth's distance from the Sun. Whether they will be seen from Earth depends, of course, on their position in the sky; the most brilliant comets usually appear in the part of the night sky above where the Sun is situated. Often, the orientation of a comet's orbit will mean that it is unfavourably placed when at its closest, and it is easy to miss the return of a faint periodic comet in this way. On other occasions, it may happen that the combined movements of the Earth and a comet will make the latter appear to stand still against the star background; alternatively, it may appear to be heading directly towards the Earth, or simply drifting away at an angle.

As the above remarks no doubt make clear, it is possible to observe comets only when they are in the part of their orbit nearest the Sun. There are two reasons for this: they do not become bright enough to be seen until they have come within a certain distance of the Sun, and their orbits may take them right into the twilight zone of the Earth's

Fig. 451 — Bright comets frequently appear in the sky as twilight is fading: Comet Hassell, April 1939.

Fig. 452 — Photograph of a faint comet, showing how difficult it is to pick it out from the starry background.

NAME	Last Observed Perihelion	Period	Peri. Dist.	Eccent.	Obs'd returns
Encke	1957	3.30	0.338	0.847	45
Grigg-Skjellerup	1957	4.90	0.856	0.704	9
Tempel II	1956	5.30	1.391	0.543	12
Neujmin II	1927	5.43	1.338	0.567	2
Tuttle-Giacobini	1951	5.49	1.117	0.641	3
Tempel-Swift	1908	5.68	1.153	0.638	4
Pons-Winnecke	1951	6.12	1.159	0.654	15
Kopff	1957	6.18	1.495	0.556	8
Forbes	1954	6.42	1.545	0.553	4
Perrine	1955	6.46	1.154	0.667	3
Schwassmann-Wachmann II	1954	6.53	2.150	0.385	5
Reinmuth II	1953	6.59	1.867	0.469	2
Giacobini-Zinner	1946	6.59	0.996	0.717	6
Daniel	1950	6.66	1.465	0.586	4
Wirtanen	1954	6.69	1.625	0.542	2
D'Arrest	1950	6.70	1.378	0.612	10
Finlay	1953	6.81	1.049	0.708	6
Brooks II	1953	6.93	1.866	0.487	9
Borelly	1954	7.01	1.448	0.604	6
Faye	1954	7.41	1.652	0.565	14
Whipple	1955	7.42	2.450	0.356	4
Ashbrook-Jackson	1955	7.51	2.324	0.394	2
Reinmuth I	1957	7.69	2.037	0.477	4
Oterma	—	7.89	3.390	0.144	—
Schaumasse	1952	8.17	1.194	0.706	5
Wolf I	1950	8.42	2.498	0.396	9
Comas-Sola	1952	8.55	1.766	0.578	4
Väisälä	1949	10.52	1.752	0.634	2
Neujmin III	1951	10.95	2.032	0.588	2
Gale	1938	10.99	1.183	0.761	2
Tuttle	1939	13.61	1.022	0.821	8
Schwassmann-Wachmann I	—	16.15	5.523	0.136	—
Neujmin I	1948	17.93	1.547	0.774	3
Crommelin	1956	27.87	0.743	0.919	5
Stephan-Oterma	1942	38.96	1.596	0.861	2
Westphal	1913	61.73	1.254	0.920	2
Brorsen-Metcalf	1919	69.06	0.485	0.971	2
Olbers	1956	69.57	1.179	0.930	3
Pons-Brooks	1953	70.88	0.774	0.955	3
Halley	1910	76.03	0.587	0.967	29
Herschel-Rigollet	1939	156.0	0.748	0.974	2
Grigg-Mellish	1907	164.3	0.923	0.969	2
LOST PERIODIC COMETS					
Brorsen	1879	5.46	0.590	0.810	5
Tempel I	1879	5.98	1.771	0.463	3
Biela	1852	6.62	0.861	0.756	6
Holmes	1906	6.86	2.122	0.412	3

atmosphere, where, of course, they are too faint to be seen.

New comets are discovered only when fulfilling the above conditions. It may be argued that this makes their discovery nothing more than a matter of luck; there are people, however, who put their luck to work by systematically sweeping over the sky every night, searching for a new comet. This method has been much improved upon in recent years by the application of photography, which offers considerable advantages over purely visual methods. There are quite a large number of workers engaged on this rather special sport of 'comet hunting', and it not infrequently occurs that two independent workers simultaneously announce the discovery of the same comet. There is also the case of the astronomer engaged on a totally different problem, who accidentally records an unexpected comet on one of his photographic plates, and afterwards identifies it only on learning its position from other sources. Comet hunting, moreover, is not confined to the professional: many discoveries have been made by assiduous amateurs using either their unaided eye or else some simple telescope which they have constructed themselves.

Because of this constant scrutiny of the skies, quite a few comets are discovered every year, though most of them are unfortunately very faint. Nevertheless, we cannot claim to discover all of these strange visitors from space: many are too far away or too faint to be seen, and others escape detection because of some unlucky combination of circumstances.

Numbers of comets. — As the above considerations indicate, the total number of comets is unknown and

Fig. 453 — USUAL APPEARANCE OF A COMET, SHOWING THE NUCLEUS, COMA AND TAIL

likely to remain so, but it is certainly large. Indeed, Kepler once remarked: 'There are as many comets in the heavens as there are fish in the sea.'

Theoretical considerations suggest that the number of comets in the Solar System must run into millions. Assuming an even distribution of perihelia, there must be about 6,000 comets contained in a sphere equal in diameter to the orbit of Neptune. How many of them do we actually observe?

Fig. 454 — COMET BROOKS IN 1911. Note the rounded head and the long, thin tail.

The most ancient observations known to science, those contained in Chinese annals, go back to the year 2369 B.C. Since then there have been 1,600 sightings, including the returns of periodic comets. Actually, this figure ought to be much higher, because there is a considerable difference between the numbers seen before and after the invention of the telescope. Before, only naked-eye comets found their way into the records, but comets as bright as this are rare, and appear only at intervals of a few years. With the application of first the telescope and then the camera, there was a rapid rise in the number of faint comets discovered — these could never be seen with the naked eye — and about a dozen are added to the list annually.

The orbital nature of many comets is not known with any great precision, either because the particular comets were observed casually by people who knew nothing of astronomy, or because the observations were made with insufficient accuracy. This lack of information applies at the present time to some of the very faint comets which are difficult to observe without powerful instruments.

By 1935, the elements of only 663 cometary apparitions had been accurately determined; according to Yamamoto's catalogue, this yields 467 separate comets. Of these, 170 have elliptical, 250 parabolic, and 47 slightly hyperbolic orbits.

This number has, of course, grown somewhat since the catalogue was compiled, as new comets are being added all the time.

Distinction must be made between those comets which have been observed more than once, i.e. periodic comets, and those which have made only one appearance. In the latter category must be placed those whose orbits have not yet been determined with sufficient accuracy, especially if their revolution period is thought to be greater than 200 years. In addition, there are comets which, although having a fairly short revolution period, have not yet returned since their first appearance, or whose return could not be observed because of unfavourable conditions. There are also cases of comets being lost, either after being thrown off course by the gravitational attraction of a large planet, or because they have actually ceased to exist, having for some reason disintegrated. (We shall return to this question later.) The term 'periodic' should not, therefore, be taken as a guarantee that a comet's return will always take place and will be visible.

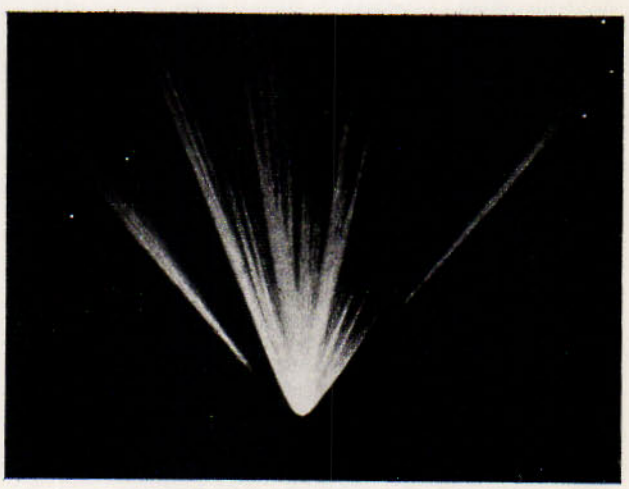


Fig. 455 — THE TAIL OF THE GREAT COMET OF 1861 LOOKED LIKE A BEAUTIFUL FAN AT ITS CLOSEST APPROACH TO THE EARTH ON 30 JUNE

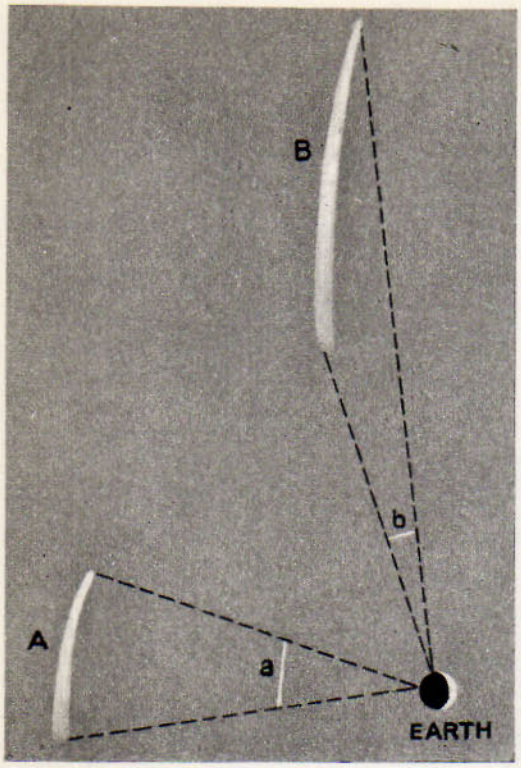


Fig. 456 — THE APPARENT LENGTH OF A COMET'S TAIL DEPENDS ON ITS ORIENTATION RELATIVE TO THE EARTH

Fig. 457 — BRIGHT CLOUDS OF GAS FREQUENTLY APPEAR TO MOVE AWAY FROM THE HEADS OF COMETS

Fig. 458 — THE TAILS OF COMETS FREQUENTLY SPLIT LENGTHWISE TO FORM SECONDARY TAILS

Lastly, it may be noted that it is very difficult to distinguish whether a comet is travelling along a parabolic orbit or along an elliptical orbit of such high eccentricity that the revolution period is of several thousand years' duration, since, in the short distance over which we can observe the object, the two types of orbit are practically identical. For this reason it is impossible to be certain exactly how many periodic comets there are.

The table on page 232 (by F. Baldet) gives the names of some of the most famous and recent periodic comets. The astronomical unit, it will be recalled, is the mean distance between the Sun and Earth. The names by which comets are known are usually those of their discoverer, although some, such as Halley's, commemorate the astronomers who worked out their orbits.

The appearance of comets. — A comet is usually shown as a large fuzzy object with a long tail. This, however, is an over-simplification and is only partly true. Although some comets conform to this typical picture, many of the fainter ones appear as nothing more than hazy spots in the telescope, with no sign of tails.

The truth is that comets are very varied both in size and shape. They possess one thing in common — their nebulous appearance — which suggests that they consist of nothing but an incredibly tenuous cloud of gas.

Broadly speaking, a typical comet consists of a bright nucleus, surrounded by a luminous cloud, the coma. The coma and nucleus together make up the head. The tail, if present, is usually a fainter streak of cloudy material joined to the head. The exact significance of this will become apparent when we discuss the cause of this phenomenon.

These distinctions between the different parts of a comet are of purely descriptive significance, and the relative sizes of these parts vary from one comet to another. For example, the nucleus is sometimes barely visible, and on occasion it may be multiple. Its appearance in such cases varies from that of a hard point of light to a brilliant patch of condensed nebulosity. Similarly, the coma is sometimes no more than a single fuzzy cloud, whilst other comas may show a remarkable structure, suggesting a tail in process of formation. The heads of different comets vary widely in appearance. Sometimes a comet appears with what is clearly an embryonic tail and the object takes the form of a slightly elongated patch of nebulosity. Others have tails which are truly fantastic in size, occasionally stretching up into the sky like the beam of a searchlight. These tails are often slightly, or even markedly, curved. It is by no means unknown for there to be small, subsidiary 'jets' projecting backwards at an angle from the coma, or for the head to be surrounded by minor nodules of milky nebulosity. This occasionally produces the effect of a splendid celestial fan, the most famous example of which was the comet with seven tails which appeared in 1744. Of a similar kind was the great comet of 1861; on the morning of 30 June the tail reached right out to the Earth.

In this connection, however, allowance must be made for the effects of perspective, because of the many different ways in which a comet may approach the Earth. The tail may be pointing directly towards us, or may be presented edge-on. Fig. 456 shows that in the first case, the tail appears to be compressed and consequently much shorter than it really is. It could, indeed, seem shorter than a smaller one seen sideways on.

When the tail of a relatively close comet is pointing towards the Earth, the effects of perspective are very pronounced: a cylindrical tail appears fan-shaped — in precisely the same way as the parallel lines of a railway track appear to converge towards the horizon. The nearer the Earth is to the end of the tail, the wider the

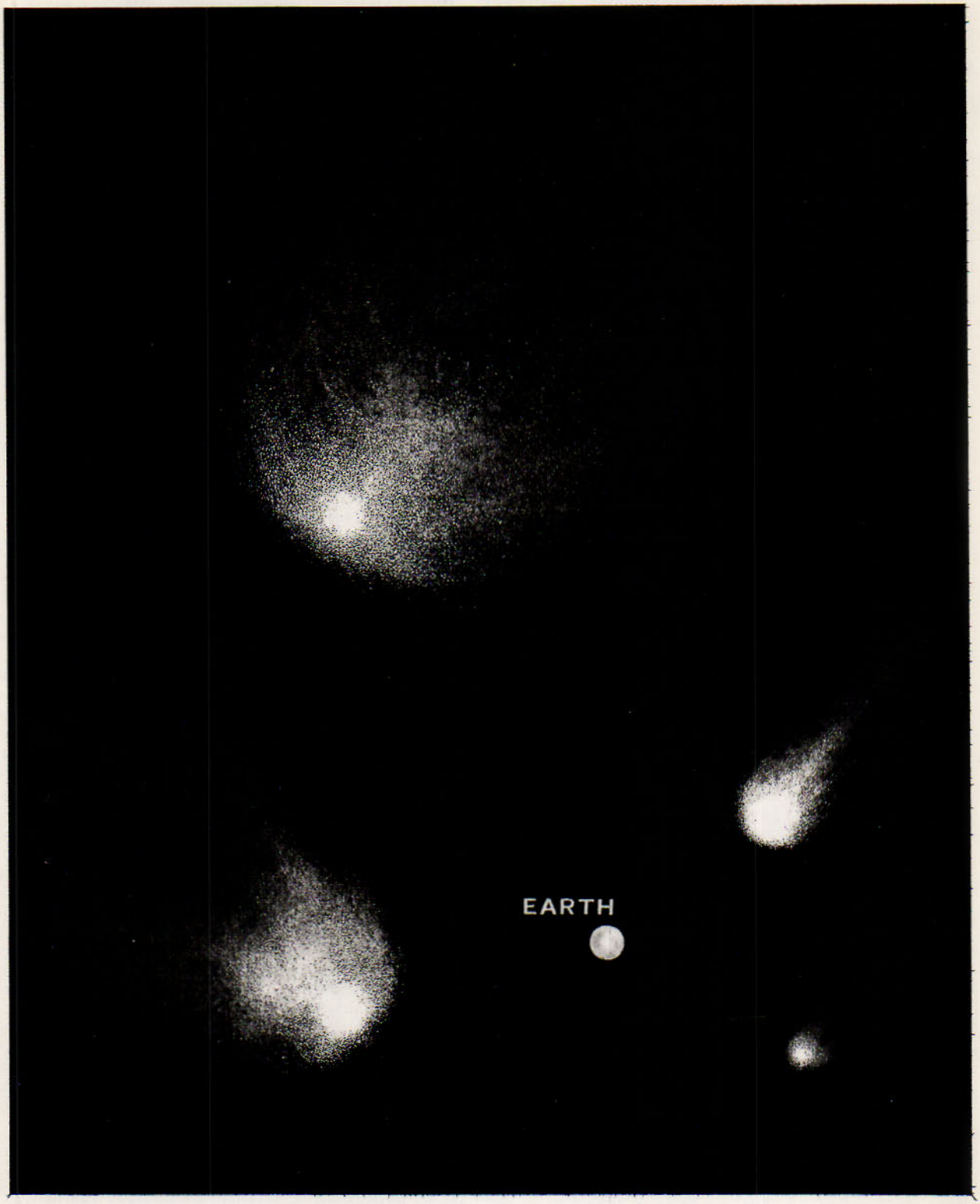


Fig. 459 — The size of some average comet heads compared with the earth

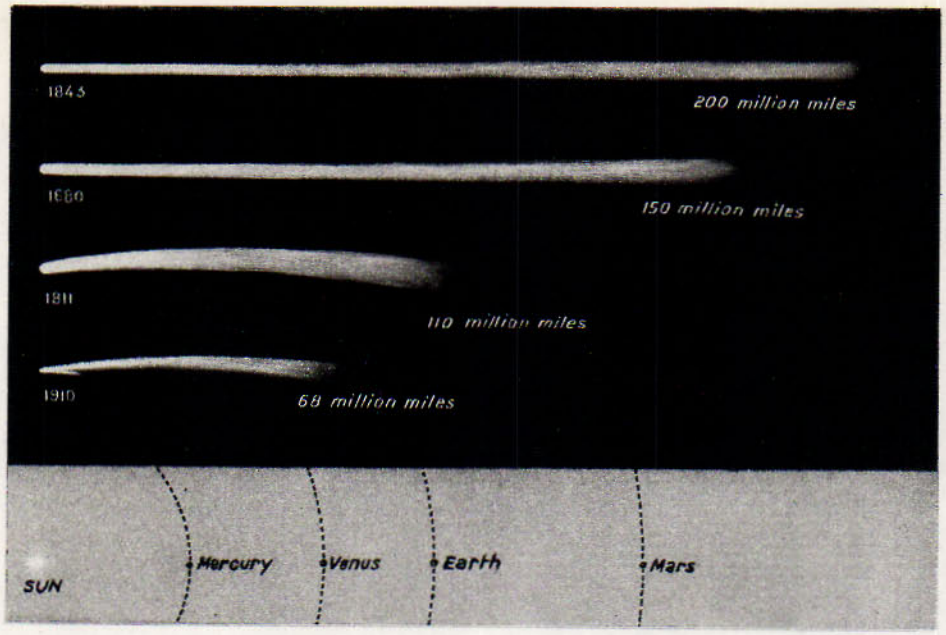


Fig. 460 — THE LENGTH OF FOUR COMET TAILS COMPARED WITH THE ORBITAL RADII OF THE INNER PLANETS

end of the fan. The same effect may alter the angle at which a multiple set of tails diverges.

It depends on perspective whether a comet's tail appears straight or curved. As will be seen later, the tails of most comets are slightly curved in the plane of their orbit. If this plane coincides with our line of sight, the curvature of the tail will not be apparent. Only an appreciable obliquity of the comet's orbital plane in relation to the Earth will show the true shape of the tail. If this principle is not recognised, it is very easy to be deceived by the frequently striking changes in the appearance of a comet.

However, this is not to say that real changes do not take place. Changes may be observed in the size and shape of the object, and in its brightness at different points on its orbit, and occasionally fragments may be seen to break off from the main body of the comet. Such happenings are best recorded photographically, since the plates are particularly sensitive to the wavelengths of light emitted by comets. The camera frequently reveals details which would escape the visual observer; indeed, it has been shown that comet tails often possess structural features unsuspected before the application of photography, and fainter extensions that the eye could never pick out.

The recognition of all these details, however, does not simplify the problem of finding a satisfactory explanation for the origin and nature of comets.

The size of comets. — No less intriguing than their appearance is the average size of comets, which, under certain circumstances, can reach remarkable proportions. As with their shape, however, there is very little uniformity in their size. Thus some comets may be large and splendid, others comparatively small and faint, although in nearly all cases they are actually large objects. Even the smallest of them are at least 10,000 miles in diameter — larger than the terrestrial globe, in fact. Usually, however, comets are constructed on a much vaster scale, and may be anything from 30,000 to 150,000 miles across, the average being about 80,000 miles. The biggest ever observed was the comet of 1811, whose enormous diameter of 1,225,000 miles was greater than the diameter of the Sun itself.

These figures give some indication of the enormous size of the heads of comets. The tails are even more tremendous, frequently reaching several million miles in length. Let us take a few examples.

The largest tail ever observed was that of the great comet of 1843, which was no less than 200,000,000 miles long; this means that it would have stretched considerably further than from the Sun to Mars. Smaller ones have exceeded the radius of the orbits of the Earth and Venus. These are illustrated in Fig. 460. Moreover, since some comet tails diverge from the head outwards, they occasionally reach several million miles in width at the broad end. Thus it is easy to understand why a large comet passing relatively close to the Earth can produce such a superb spectacle in our skies.

It must always be borne in mind, however, that whether or not a grand spectacle occurs depends on the circumstances mentioned earlier; in most cases it is not the really immense comets but the smaller ones which look biggest in the sky. The comet of 1843, for example, with its 200,000,000-mile tail, covered an angular distance in the sky of 68°, whereas the comet of 1861 subtended no less than 118°, although only one-third as large.

However, the figures given in this section must only be taken as fairly close approximations to the real dimensions. Exact measurement is a matter of extreme delicacy, since observations of a fuzzy and possibly faint object are subject to the atmospheric conditions obtaining, the power and quality of the instrument in use and also the means employed (visual or photographic) in studying the object.

The structure of comets. — Nothing in the appearance of comets, with the exception of the nucleus, would suggest an assemblage of material in the solid state; on the contrary, everything seems to indicate that these vast nebulous masses are extremely transparent, similar, for example, to an ultra-tenuous gas. That they are constituted in this way — which is suggested by simple visual observation, since even quite faint stars will

Fig. 461 — Through the telescope, a comet passing close to the Earth can be seen as a tiny nucleus accompanied by little jets of gas which appear as tufts in the coma.

shine through a comet passing in front of them with no appreciable diminution in brightness — is confirmed by another approach. Comets do not exercise any appreciable perturbing influence on the planets. For example, in 1886 a comet passed through Jupiter's system of satellites without causing the least disturbance. In other cases, it is possible to make a calculation, which, if not exact, is indicative: a comet may be several million miles in diameter, but its mass represents less than a thousand million metric tons weight — that is, it weighs much less than a cubic mile of water.

Although these estimates should not be taken as anything more than an attempt to give some idea of what we are dealing with, they do at least show that comets are objects of extraordinary tenuity. Their mean density is actually much less than that of the best vacuum which can be produced in the laboratory, and this fact becomes even more striking when we remember that most of a comet's mass is concentrated in the nucleus and coma, which are relatively small. The tremendous rarefaction of this gas is more vividly illustrated when expressed in terms of pressure: in the region of the nucleus, where it is densest, it would exert a pressure of 1/10,000,000,000 of a millimetre of mercury. Under these conditions, 10,000,000,000,000 times less dense than air, a molecule of gas can travel a mean distance of 600 miles without colliding with one of its 'neighbours', compared with 1/10,000 of a millimetre in air at sea level pressure. In the tail, which is even more tenuous, the molecules can travel up to a million miles without colliding, which makes them almost completely independent of one another. Under these conditions, they can remain for long periods in states which are highly abnormal judged by terrestrial standards.

The precise nature of the nucleus remains something of a mystery. As we have said, it is sometimes difficult to be certain as to what the nucleus actually is, since it is frequently lost in the surrounding cloud of gas. Visual observations, however, with very powerful telescopes, indicate that it is an extremely minute object, contrary to what might be thought from its appearance under normal conditions. Indeed, F. Baldet has suggested that it may not be much more than a few hundred yards across. Opinions differ as to whether it is a single body or a collection of small fragments, cohering under their mutual gravitation. The second suggestion is possibly supported by the fact that — as will be discussed later — comets are fragile objects, liable to disintegrate under certain conditions. Moreover, it is consistent with the fact that on the rare occasions when a comet has passed in front of the Sun no sign of the nucleus has ever been observed.

The rest of the formation is thought to be composed of extremely rarefied gases which are given off under the influence of solar radiation. Although we do not yet know the exact mechanism whereby this takes place, the effect seems clearly to be present.

The general appearance of a comet passes through an extraordinary sequence of changes during its passage round the Sun. The head contracts sharply, reaching its minimum size a month or so after perihelion. Simultaneously, however, the tail often lengthens out to a tremendous size. These two opposite trends are not incompatible as might at first appear. They can be explained by recourse to a remarkable fact about the fundamental nature of comet tails.

Our first impression on seeing the picture of a comet is that the tail is following the head, rather like a flag

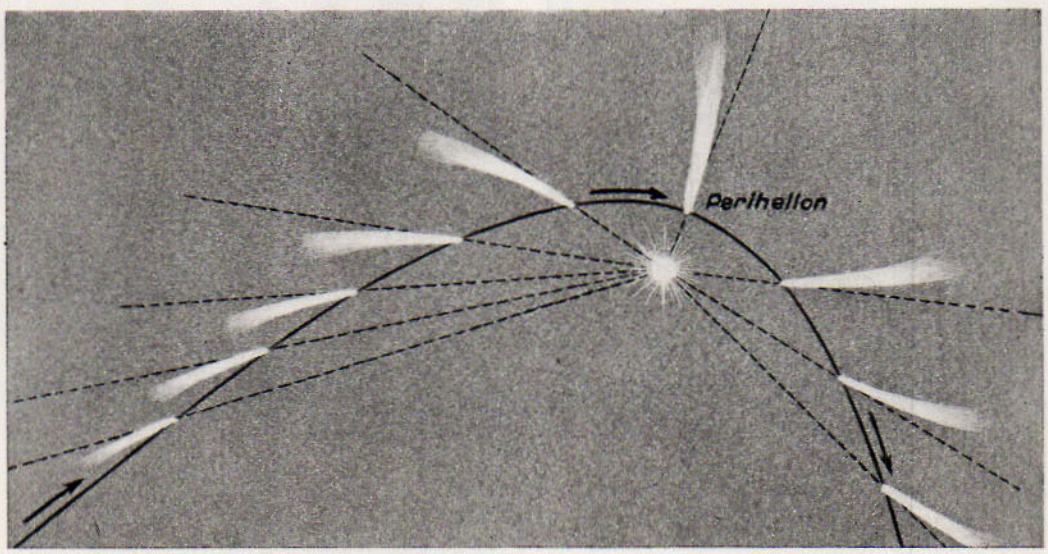


Fig. 462 — A COMET'S TAIL ALWAYS POINTS AWAY FROM THE SUN

19 September 1908

30 September 1908

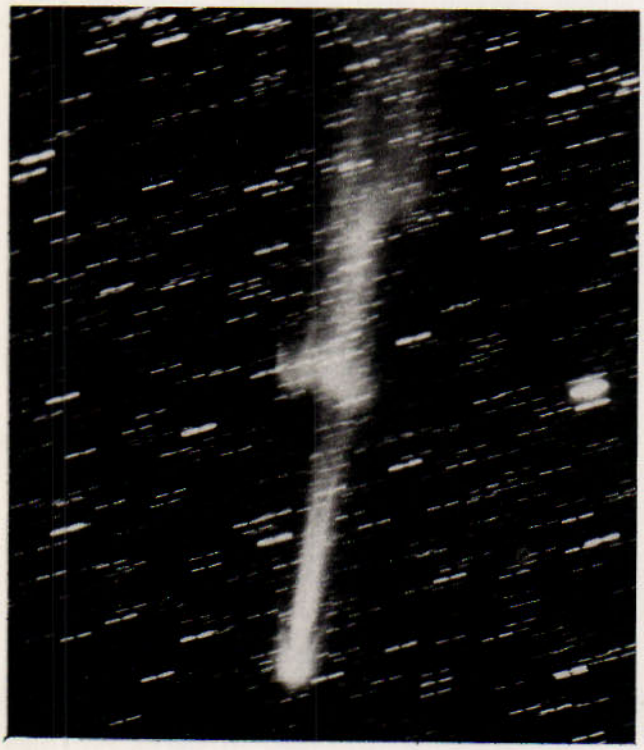
1 October 1908

4 October 1908

Fig. 463 — A REMARKABLE SERIES OF CHANGES IN THE APPEARANCE OF THE TAIL OF COMET MOREHOUSE, 1908. Photographs by E. E. Barnard (Yerkes Observatory) and F. Quénisset (Juvisy Observatory). *(Palais de la Découverte, Paris)*

14 October 1908

15 October 1908



16 October 1908

17 October 1908

Fig. 463 — A REMARKABLE SERIES OF CHANGES IN THE APPEARANCE OF THE TAIL OF COMET MOREHOUSE, 1908. Photographs by E. E. Barnard (Yerkes Observatory) and F. Quénisset (Juvisy Observatory). *(Palais de la Découverte, Paris)*

billowing out behind a ship. This idea is quite mistaken. The tail of a comet always points away from the Sun — as if the flag were being blown at right angles to the ship's motion by a stiff cross-breeze — and is therefore constantly changing direction. It is the head which describes the elliptical or parabolic orbit; the tail follows the head before perihelion passage and precedes it afterwards. When nearest the Sun, the comet is moving very rapidly and so the tail swings round at great speed, like the rotating beam of a lighthouse.

This orientation of comet tails away from the Sun has been known for a long time; comets were considered by some ancient philosophers to be false images formed in the sky by reflection of sunlight. With the increasing accuracy of observations, scientists looked for some more rational explanation; in 1554, Cardan suggested that comets behave rather in the manner of lenses, refracting the light of the Sun into a cone behind them, thus giving the appearance of a tail.

However, it was Kepler who, in 1617, attempted the first truly rational explanation. He considered comet tails to be due to some sort of repulsive force emanating from the Sun and driving the gaseous matter in the head out into a long streamer. This could be compared with the action of blowing a trail of smoke from a match. In terms of modern physics, the Sun's 'breath' is the actual pressure of its radiation on the molecules of dust and gas of which comets are formed. In fact, a ray of light behaves rather like a stream of tiny corpuscles, and when it strikes some obstacle, the latter receives a certain impulsion. This is demonstrated by a simple, classic experiment: if an intense pencil of light is directed upon a slender thread of lycopodium powder falling in an hourglass, the powder is seen to deviate by a small amount from the vertical, as if swayed by a gentle breath of wind. Radiation pressure affects all objects, but its effect becomes dominant over the mutual gravitation of two objects towards each other when the size becomes very small (below a thousandth of a millimetre, for example, in the vicinity of the Sun). The reason for this is that the surface exposed to the action of light becomes larger in relation to the volume — and therefore the mass — of the particle, which then begins to move under the radiation pressure. In the case of a comet, particles are driven away in the opposite direction to the Sun to form the tail.

The heads and tails of comets, however, are composed not of fine dust but of isolated molecules of gas, and the mechanism involved is a slightly more subtle one involving the fundamental process whereby packets of light energy (photons) are absorbed by the molecules. (This subject is discussed at greater length in Chapter 20.) The process of absorption includes an impetus — as in the example above — which projects the molecule in the opposite direction to that from which the photon has arrived.

In the case of isolated molecules, this process can only take place in certain frequencies or wavelengths of light, which are, as it were, 'tuned' to the vibrations of the molecule. For this reason it would be more accurate to speak of selective radiation pressure, or resonance pressure.

Once accepted, this mechanism explains why the tail develops at the expense of the head, and also why the head shrinks and the tail grows in the vicinity of the Sun.

Considering only the principle involved in its formation, it might be thought that a comet's tail ought to be straight. It has already been pointed out, however, that comets' tails are usually curved. The reason for this lies in the various motions involved: the curvature of the tail is the resultant of these motions since it is made up of molecules projected backwards at different points along the comet's path. This may be illustrated by a simple experiment. Take a hosepipe which is spurting water under fairly high pressure; if you swing it round in a wide arc the jet of water will be noticeably curved, in the same way as a comet's tail.

This general account of how comet's tails are formed is confirmed by the results of spectroscopic analysis, and by both visual and photographic observations. In the head, jets and spikes are occasionally seen pointing backwards or fanning out slightly after the fashion of the tail. This sort of structure may be prolonged into the heart of the tail, becoming more complex and acquiring long filaments and denser patches, all in a state of turbulence. These may recede from the head with velocities often exceeding 10 miles per second. The most striking example of this was furnished by Comet Morehouse in 1908, whose tail on several occasions appeared not only to dissolve into fragments but also to separate completely from the head (see Fig. 463, pp. 238–239). This suggests some sort of turbulence in the constituent parts of the comet as well as an uneven emission of gas by the head.

The light given out by comets has a dual origin. Like the planets, they are illuminated by the Sun and reflect its light. For the most part, however, the luminosity which we observe is due to a glow produced in the gas of the comet principally by the ultra-violet radiation of the Sun. This explains why the brightness of comets, which is a function of their distance, is not very great until they reach the outer regions of the Solar System.

The spectroscope reveals that the heads of comets are principally composed of neutral molecules of carbon (C_2). neutral molecules of cyanogen (CN) and hydrocarbons (CH), whilst groups of OH and NH have been observed from time to time. To these diatomic molecules, which are incomplete from the point of view of terrestrial chemistry and which can only exist under these conditions by virtue of the extreme tenuity of the gas, may be added the triatomic molecule CH_2, which has been identified by the physicist Herzberg in the vicinity of the nucleus. Herzberg has been able to reproduce its spectrum in the laboratory, using a special technique for exciting methane electrically.

When the comet approaches the Sun, the temperature of the nucleus increases, and it begins to emit sodium vapour, whose characteristic yellow radiation is sometimes sufficiently intense to be photographed in full daylight, rendering the comet quite plain even at midday (Comet Skjellerup, 1927) even when only a few degrees from the Sun.

Finally, comets' tails contain molecules of ionised carbon monoxide (CO^+) and ionised nitrogen (N_2^+); such ionisation (the stripping off of one or more planetary electrons) is closely connected with the process whereby the tail is formed from molecules violently expelled from the head under the influence of ultraviolet radiation from the Sun.

Of course, there is still much to be learned about these strange wandering objects, whose existence we have good reason to think much more ephemeral than that of either the stars or planets. It is inconceivable, for example, that the emission of gas from the head to form the tail could continue indenfinitely, particularly in the case of the periodic comets which frequently feel the full intensity of the Sun's radiation. Moreover, this is supported by the observational evidence, which shows, sometimes quite clearly, that the small solid mass constituting the nucleus is liable to disintegrate, gradually strewing fragments of débris all round the comet's orbit.

Thus it can happen that a periodic comet fails to appear when due because it has disintegrated — as in the famous case of Biela's Comet — and then completely

dispersed during succeeding revolutions. This question of the breaking up of comets is connected with the study of meteors or shooting stars, and will be discussed at length in the next chapter.

Comets and the Earth. — Although the stars in general have always held men fascinated in spite of our long ignorance as to their true nature, bright comets enjoy the singular distinction of being the cynosure of all eyes whenever they appear, giving birth to the most fantastic ideas in the minds of men. As we have said, everything conspires to make them stand out as something special: their appearance, so different from that of any other heavenly body, their habit of suddenly springing into view, and the apparently unusual paths that they follow across the celestial sphere.

Both the characteristic shape of comets and their variety of shapes have been known since the earliest times. On the other hand, men have always tended to view them with terror, especially in the Middle Ages, and even in modern times in primitive societies. Imagination, it seems, lends wings to the flight of reason. Thus we find accounts which, whilst often naive or even incorrect, do capture the essential spirit of comets, as seen through these men's eyes. The historian Nicetus wrote in 1182: '. . . a comet appeared in the heavens like a twisting serpent; now writhing and coiling back upon itself, now it would terrify the people with its gaping mouth; as if lusting for human blood, it seemed about to slake its thirst. . .' And the famous surgeon Ambroise Paré outvies Nicetus, in his account of the great comet of 1528: 'So horrible was it, so terrible, so great a fright did it engender in the populace, that some died of fear, others fell sick . . . this comet was the colour of blood; at its extremity we saw the shape of an arm holding a great sword as if about to strike us down. At the end of the blade there were three stars. On either side of the rays of this comet were seen great number of axes, knives, bloody swords, amongst which were great number of hideous human faces, with beards and hair all awry . . .'

Let us not pursue the matter further. These two passages clearly show the terror that comets could inspire, and the way in which imagination, prompted by ignorance and superstition, could run wild. The mysterious appearance of a comet was thought to be the presage of tragedy or catastrophe; we all know Shakespeare's famous lines from *Julius Caesar*:

When beggars die, then are no comets seen:
The heavens themselves blaze forth the death of princes.

It must be admitted, to our shame, that in spite of the tremendous advance in scientific knowledge during the last few centuries, these old fears linger on in one form or another and can be met with even today.

From being 'signs of heavenly wrath', comets were transformed by the work of Newton and Halley into astronomical bodies obeying the inviolable laws of universal gravitation. As the study of their orbits progressed, it became evident that some cross the Earth's orbit, and that there was a consequent risk of an encounter between our world and a comet.

It was first imagined that the result of such a meeting would be, to say the least, considerable, if not calamitous. Thus we find an account of such an event in Laplace's immortal work, *Exposition du système du Monde*: 'the axis and rotational movement (of the Earth) changed, the seas forsaking their age-old positions and rushing towards the new equator; most of the human race and the beasts of the field were drowned in this universal deluge, or destroyed by the violent shock imparted to the terrestrial globe, entire species annihilated, every monument to human endeavour overthrown . . .'

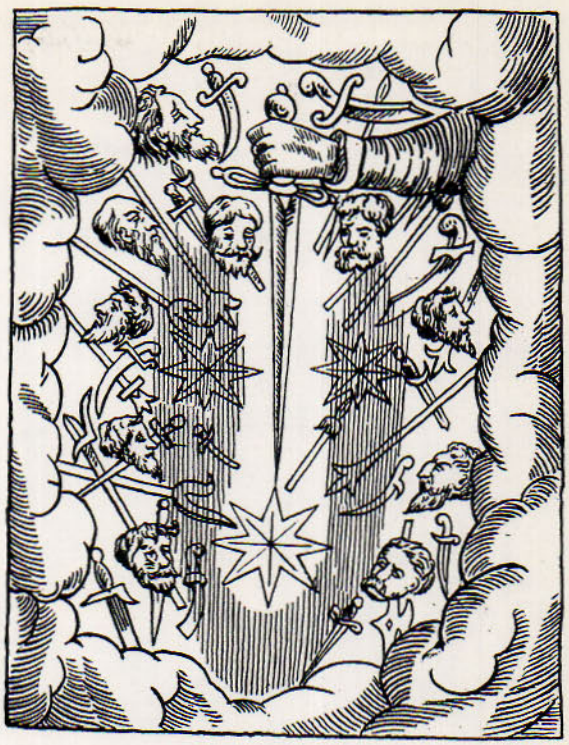


Fig. 464 — THE COMET OF 1528 DESCRIBED BY AMBROISE PARÉ, FROM A CONTEMPORARY ENGRAVING

In popular thought, this picture has attached itself to the idea of the 'end of the world'. It should, however, be noted that Laplace specifically stated that all this would take place only if the mass of the comet were about equal to that of the Earth. Now, since by far the majority of comets have only a tiny fraction of the Earth's mass, a collision with one of them would produce only local upsets at the most. Modern knowledge, indeed, confirms the famous mathematician's reservations on this point.

In spite of this, however, these questions are of considerable interest and are worth a more detailed treatment in the light of modern knowledge. Various phenomena, of which more will be said, and some more direct evidence of a different kind, furnish a useful guide.

First, however, let us examine the question of a collision between the Earth and a comet. What we are actually considering here, of course, is an encounter between the terrestrial globe and the head or, more specifically, the nucleus of a comet. An actual collision would take place only if the two orbits intersected exactly, and if the Earth and comet arrived at this point simultaneously, like two vehicles at a cross-roads. Since the nucleus is surrounded by the vast cloud of gaseous material which goes to make up the head, the Earth would pass through this before actually colliding with the nucleus. The perfect timing and positioning required make the possibility of collision extremely slight.

The probability increases if we consider collision with the coma only, since the latter is so huge. The Earth might encounter it — or more correctly, pass through a part of it — even if the two orbits did not cross exactly. It is only necessary that the distance separating them should be less than the radius of the coma.

It is easy to see that the second possibility is the more likely. Let us see if the facts confirm this.

Ever since man turned his gaze towards the heavens,

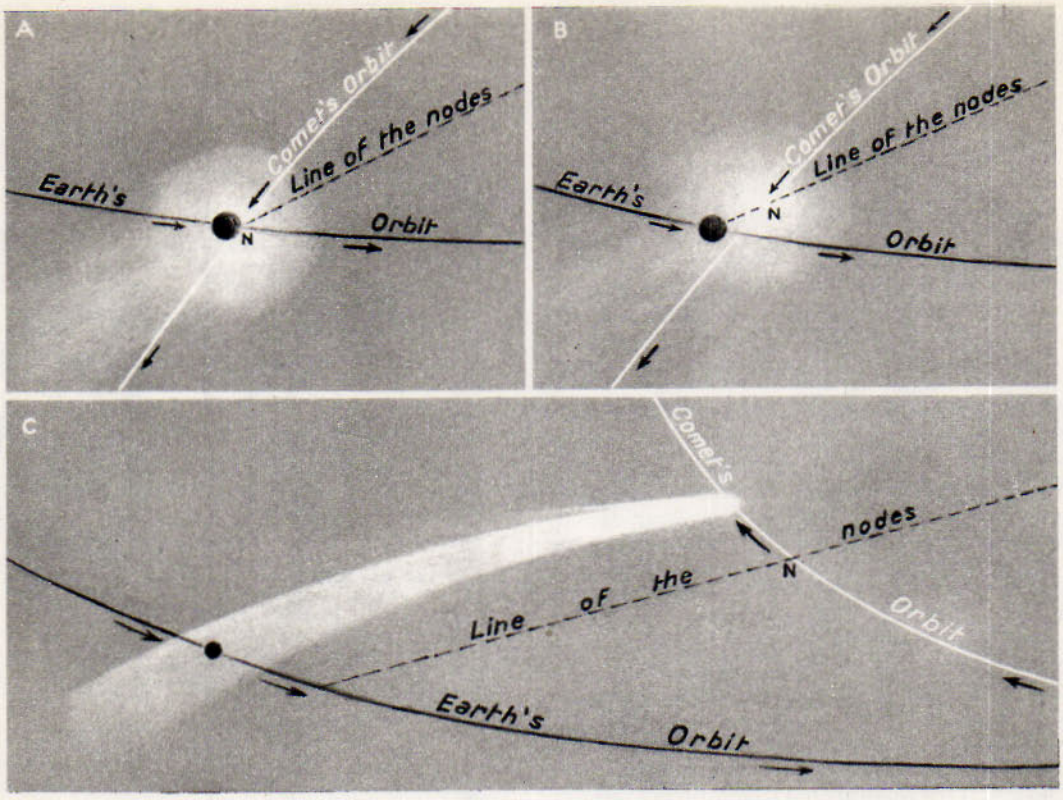


Fig. 465 — THREE DIFFERENT WAYS IN WHICH THE EARTH COULD ENCOUNTER A COMET. A, the nucleus collides with the terrestrial globe. B, the nucleus misses the Earth, which passes through the coma. C, the Earth passes through the tail.

no comet has made a direct hit on the Earth, though twice since the middle of last century — on 30 June 1861 and again on 19 May 1910 — the Earth has passed through the end of a comet's tail. In spite of this, however, let us see what outward signs each different type of encounter would produce.

As regards a direct collision with the nucleus, we do not have to rely on guesswork alone. Although the prospect is not entirely reassuring, it need not give rise to exaggerated fears, and if such fears are occasionally voiced nowadays, it is because the real nature of comets has been misunderstood and that they have been credited with a solidity which they do not possess.

The first thing to remember is that the nucleus of a comet — whether a solid lump or a cluster of fragments — is a very tiny object, perhaps only a few hundred yards across. Compared with the Earth, this is like an insect trying to knock down a locomotive. We can be certain that the locomotive will suffer no harm, but can the same thing be said of the Earth? Obviously, it would not be the end of our world, whose path in space would not even wobble, but the part of the surface which received the impact of the nucleus would be severely devastated, and this devastation would probably extend over a fairly wide area because of shock waves in the air and ground. By normal standards, we should be dealing with a tremendous explosion due to the sudden release of the kinetic energy of an object weighing several million tons travelling at a velocity of anything up to several hundred miles per second.

It is possible to obtain some idea of what this implies by working up from something of the same type which occurs, fortunately, on a much smaller scale. In this connection, meteorites, which will be discussed in the next chapter, provide a useful basis for comparison.

Apart from the actual impact, there may be effects, possibly of an electrical nature, due to the passage of the Earth's atmosphere through the cloud of gas forming the comet's head, which is in a state not normally encountered on Earth. However, the importance of this can be exaggerated, since, as we saw earlier, these gases are extremely tenuous. This is confirmed by the fact that nothing untoward happened on either of the two recent occasions when the Earth passed through a comet's tail, the comets concerned being the great comet of 1861 and Halley's Comet of 1910.

The first of these went by quite unnoticed except by people who were expecting it. This indicates how harmless it was. As for the second, many readers will have seen for themselves that nothing tragic took place, except perhaps for the exaggerated fears manifested by people who believed sensational pronouncements made in the Press. The public at large saw very little, although the initiated did observe a few items of interest: the night-sky was abnormally bright, as if Earth had plunged into a luminous mist. Moreover, in 1910, many observers, especially at the Pic du Midi, noticed diffraction effects in the form of haloes round the Sun and Moon. These could have been caused by a tenuous cloud of minute cosmic particles, of the order of a thousandth of a millimetre in diameter. At the same time, no electrical or magnetic disturbances were recorded with certainty. It appears, therefore, that if this encounter produced any effects on our planet at all, they were confined to the upper reaches of the atmosphere.

It should not be forgotten that many reservations have been expressed about the exact circumstances under which the Earth is supposed to have passed through the tails of these comets.

In theory the question is simple: the known movements

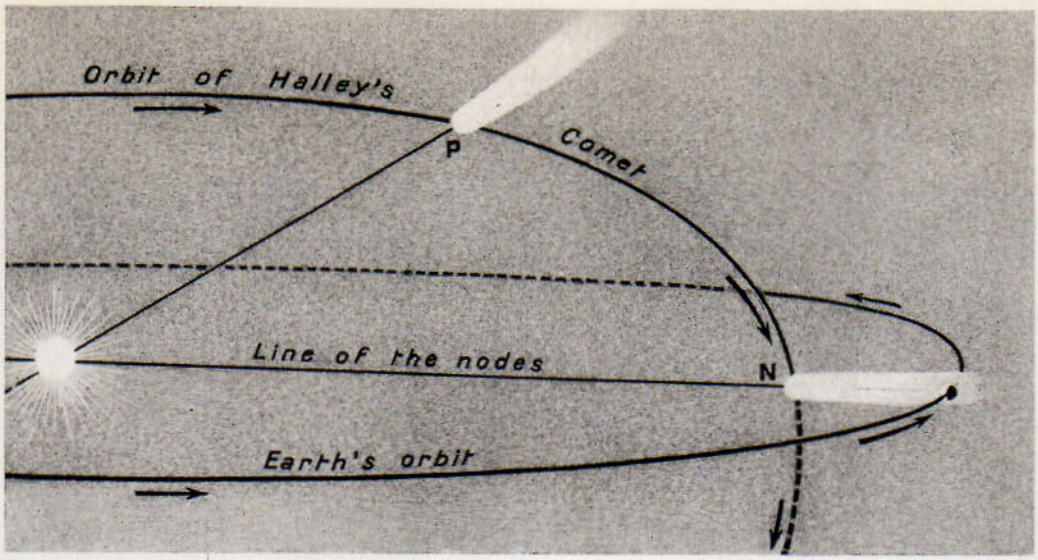


Fig. 466 — THE EARTH'S PASSAGE THROUGH THE TAIL OF HALLEY'S COMET, AS IT WAS CALCULATED TO OCCUR ON 19 MAY 1910

Fig. 467 — COMET AREND-ROLAND, 24 APRIL 1957
(Photo, B. Manning)

of the two orbits enable us clearly to define their mutual interaction. In practice, the matter is very complex, since the calculations are laborious and cannot be completely accurate for want of certain information. This means that although the orbital motion of the nucleus can be defined correctly, it is impossible to predict the exact orientation, size and shape of the tail. Just how inaccurate it is possible to be can be realised when we remember the rapid variations that have been observed in the tails of many comets.

Let us take the case of Halley's Comet. Fig. 466 shows how the Earth was expected to plunge through the comet's tail on 19 May 1910, at which time our planet should have been crossing the prolongation of the line joining the Sun to the nucleus. Now the observations made at the time (which were not all in agreement because of the comet's faintness and lack of definite outline) revealed a different state of affairs: the tail, deviating slightly from its theoretical position, was seen to be split into two unequal streamers. This fact alone raises the question of what really did happen: some maintain that the Earth passed through the smaller of the two parts, missing the larger.

Thus, if this is true, the Earth did not well and truly pass through the tail of Halley's Comet. Others, citing the absence of visual effects, contend that the Earth's magnetic field exerted a sort of screening function, and that the cometary matter was diverted round us without coming into contact at all. This claim has not been substantiated.

For this uncertainty we must blame the actual nature of comets, which are such extremely diffuse objects that it is difficult to be sure where the comet ends and space begins. At any rate, since it seems to be established that the Earth has made some sort of contact with these nebulous visitors without suffering any ill-effects, they can be judged as fairly harmless for all practical purposes. Indeed, far from dreading them, we should wish for more close contacts, since it is only in this way that astronomers will solve the many problems still attached to these mysterious objects.

Fig. 468 — A FIREBALL

Fig. 469 — A 'SHOOTING STAR' IS NOT A STAR WHICH BECOMES DETACHED FROM THE VAULT OF HEAVEN AND FALLS TO THE EARTH

CHAPTER 9

METEORS AND METEORITES

The objects commonly known as shooting stars, meteors, fireballs or bolides are familiar to everyone; equally well known is the occasional descent from the skies of lumps of stone or metal, to which the names meteorites or aerolites have been given.

These sudden and often magnificent spectacles, as well as the unexpected arrival of projectiles which may cause considerable damage, were for a long time remarked upon rather than studied. But the frequency of their appearances, their unusual numbers at certain times of the year (when they appear to radiate from definite regions of the sky), and the fall of meteorites all suggested the need for systematic observation and increasingly thorough-going investigation. As a result, some complex problems were uncovered, and these were of such interest that this particular branch of astronomy has developed rapidly in recent times.

Speaking in general terms, we are here concerned with bodies of various sizes (though they are all relatively small) which circulate throughout space in every direction and which, on encountering the Earth, have first to pass through its atmosphere. In the course of this passage, which is made at very high speeds, they are heated to a brilliant incandescence: hence their characteristic appearance of ephemeral streaks of light or of glittering stars which leave behind them a train reminiscent of that of a Guy Fawkes rocket. Thus these cosmic bodies cannot be seen so long as they remain outside the limits of the Earth's atmosphere, and become visible to us only when they penetrate it — whether or not they ultimately succeed in reaching the ground.

They may take the form of mere flashes of light or of 'stones falling from the skies', but in either case their appearance and behaviour has to be described, and many of the records of such observations tend to be confusing since different terms are often used for the same thing. Thus a brilliant shooting star may be described as a small fireball, or bolide, while the term 'meteor' has been applied indifferently to shooting stars and fireballs on the one hand, and, on the other, to those solid bodies which survive their passage through the atmosphere and which are also known as meteorites or aerolites.

Since such a diversity of terminology can only detract from the documentary value of the observational records, it is desirable that the terms should be more precisely defined; the following convention has accordingly been adopted. The term meteorite is confined to those bodies that reach the Earth's surface and also to such bodies while they are still circulating in space. The name meteor is reserved for the observable phenomena occasioned by the passage of a meteorite through the atmosphere. Thus the terms shooting star, fireball, bolide, etc., disappear from the scientific vocabulary, though they still have their picturesque and literary uses. Such phrases as 'a shower

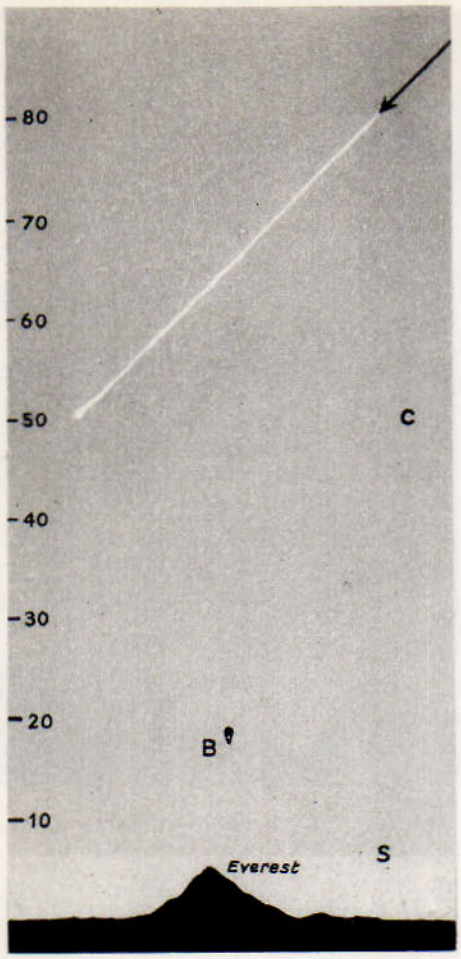


Fig. 470 — MEAN ALTITUDES (IN MILES) AT WHICH METEORS APPEAR AND DISAPPEAR. C, upper limit of the atmospheric layer whose illumination causes twilight; B, greatest altitude reached by a manned balloon; S, beginning of the stratosphere.

of shooting stars' or 'an immense fireball passed overhead, leaving a brilliant train and finally exploding' will always have their evocative value.

The adopted nomenclature may appear rather summary, in view of the diversity of the observed phenomena. But we do not yet know to what extent this diversity springs from mere differences of degree or from intrinsic differences among these celestial bodies.

The heights at which meteors appear. — The fleeting and fast-moving streak of light that, so far as appearances go, constitutes a typical meteor, occurs at a considerable height above the Earth's surface, in the extremely tenuous upper levels of its atmosphere; at this point it would be well to recollect what has been ascertained regarding the depth and structure of our atmosphere. To be precise, a meteor's incandescence is caused by its passage through the atmosphere, and occurs only when it has penetrated to a sufficient depth: when it is completely burnt out it is still far above the Earth's surface.

The innumerable measurements that have been made of the heights of meteors have shown that, although exceptionally bright ones may first become visible at some 100 miles, the average height of appearance is about 70 miles and that of disappearance between 35 and 40 miles. The length of their paths through the atmosphere (which may be more or less oblique) is therefore of the order of 50—100 miles. The fainter meteors are somewhat more restricted, with an average path length of about 35—40 miles, between an altitude of 60 and 50 miles.

These figures apply to particles of cosmic origin, whose passage through the atmosphere is revealed to us only by their incandescence — those, in other words, which for reasons which will be discussed later do not succeed in reaching the Earth's surface.

It is at once obvious from the lengths of the observed paths and the speed with which they are covered, that very high velocities must be involved. Typically, these are to be measured in tens of miles per second. Observations designed to provide more accurate data regarding meteor velocities have led to important conclusions concerning the origins of these countless inhabitants of space.

Appearance and brightness of meteors. — A meteor usually appears as a more or less bright streak of light; the reason for this is to be sought in its great velocity and in the phenomenon known as the persistence of vision. In some cases, however, this optical illusion is not the only factor involved: many meteors do in fact leave behind them a train which remains visible for some time after the extinction of the meteor itself, and which only slowly fades into invisibility; such meteors, in other words, appear to leave a phosphorescent trail in their wake. A proof of the objectivity of this phenomenon is provided by observations of persistent trains (lasting sometimes several minutes or even longer) which gradually drift and

Fig. 471 — FIREBALL LEAVING A PERSISTENT TRAIN, which becomes progressively deformed. (A. Hamon, 9 October 1933)

change their shape, just as clouds do, as a result of the movements of the atmosphere.

Such effects are particularly noticeable in the case of the very brilliant meteors commonly known as bolides or fireballs. These are no longer mere moving points of light but, as the name suggests, balls of fire, whose angular size at times appears to equal that of the Moon as seen with the naked eye. They are sometimes white, sometimes red, yellow, green or blue, and are so brilliant that in a number of instances their appearance has turned night into day; on other occasions they have been easily seen in broad daylight. It has often been observed, moreover, that they emit bright specks of light, reminiscent of the sparks that fly off a knife-grinder's wheel; these ejected 'sparks' sometimes remain visible for several minutes along the whole length of the luminous train.

The initial cause of a meteor's incandescence is the high velocity with which it plunges into our atmosphere. When we come to discuss the problems connected with the motions of bodies in space, we shall learn that in the Earth's vicinity such a body will have a velocity of about 25 m.p.s. The Earth itself is moving round its orbit with a speed of 18.5 m.p.s. If the directions of the two motions — that of the Earth and that of the meteor — are opposite, their velocities will be added to one another, so that the meteor will hurl itself into our atmosphere at about 45 m.p.s.; even in the opposite case, when the Earth and the meteor are moving in the same direction, its velocity relative to the atmosphere will be about 6 m.p.s. If its path is inclined, more or less obliquely, to the Earth's orbit, then its relative velocity will lie somewhere between these two limits. Finally it should be mentioned that, in fact, many of these cosmic fragments are travelling through space with velocities even greater than 25 m.p.s.

The incandescence that results from such enormous velocities is to be explained as follows. Despite the extreme tenuity of our atmosphere at great altitudes it nevertheless offers, under these conditions, a considerable resistance to the penetration of a meteor; the latter is therefore heated to melting point, and beyond, by the friction set up between its surface and the molecules of the atmosphere. At the same time, molecules escape from its surface, and it is the collisions of these molecules with those of the atmosphere that give rise to the radiation which we see. Laboratory experiments by the French engineers Muraour and Michel Lévy have shown that to this effect must be added the radiation that is actually produced ahead of the meteor, resulting from the very high temperature of the shock wave. It is extremely difficult to distinguish between the individual rôles of these several agencies with any degree of certainty, especially as the relative importance of each may easily vary from point to point along the meteor's path — this, at any rate, is suggested by the marked variations of intensity (even complete breaks) along a meteor's luminous path that are frequently observed.

Taking into account the total radiation emited by a meteor, it has been calculated that one as bright as the brightest stars must be equivalent to a 12,000-candle-power lamp viewed from a distance of 60 miles; yet the source of this light is a cosmic particle weighing only a few hundredths of a gram. A meteor whose brightness is comparable with the full moon would weigh about 10 lb.

This explanation assumes that deceleration due to the resistance of the air is capable of producing enough energy to vaporise solid bodies of this size before they reach the ground; and in this way the instantaneous disappearance of a meteor, while still at a great altitude, is explained. Thus only bodies of exceptional size — or, rather, their fragments — are capable of penetrating to the Earth's surface; when they do, their impact is always to some extent impressive.

Fig. 472 — A METEOR RECORDED BY A FIXED CAMERA

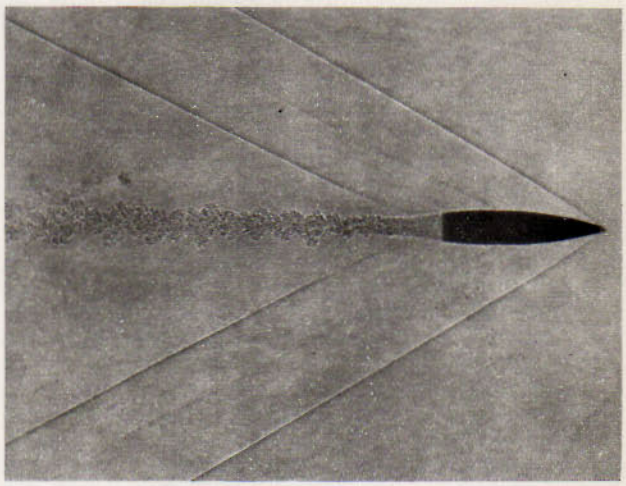


Fig. 473 — SHOCK WAVES PRODUCED BY A RIFLE BULLET TRAVELLING AT 2,190 FEET PER SECOND. *(Photo, Stroborama, Séguin Brothers; Gen. Libessart's technique)*

We have just mentioned fragmentation. Some fireballs do in fact explode like bombs, probably because of the forces acting upon the lump of meteoric material which are generated by its extremely rapid passage through the atmosphere — intense and virtually instantaneous heating of its surface, and enormous pressures which tend to crush it to pieces. The constitution of some meteorites is such that they are unable to withstand these, and they disrupt explosively.

The high surface temperature of a meteorite is shown by its fused or vitrified appearance, and also by the fact that immediately after its fall — if one has the good fortune to be near the spot — it will be found to be too hot to handle. It has also been established that this heat has not had time to be conducted into the interior of the meteorite, for if it is shattered on striking the ground, or is broken artificially, it is found that its interior is glacially cold.

In many cases the appearance of a really 'large' meteor is accompanied by a series of detonations, resembling

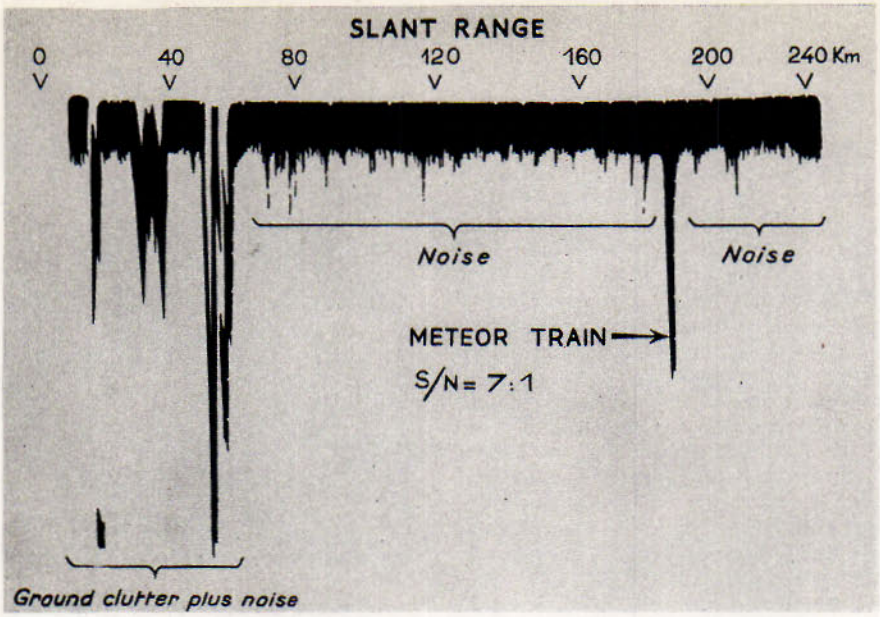


Fig. 474 — Appearance on the screen of a radar oscillograph of an echo from a meteor train. (*Sky and Telescope*, March 1947)

a long rumble of thunder. These noises are produced by the shock waves set up by its vertiginous speed, and must not be confused with the crash caused by its explosion. However, it is at times difficult to distinguish between the two, owing to the widely differing distances of meteors and to the successive waves produced all along their path; the successive passages of these often cause a detonation whose direction it is impossible to specify.

The continued luminescence of the train, after the passage of the meteorite itself, is at first sight puzzling. One cannot suppose that the particles torn from the meteoric mass will remain incandescent solely on account of their high temperature: in order to maintain such a state for some time the débris would have to be so massive that it could not remain suspended in the atmosphere. Residual débris is to be found along the path, certainly, but as we shall see this is so tenuous that it would cool immediately. One must therefore appeal to another type of phenomenon — the excitation of the molecules of the upper atmosphere by the shock of the meteorite's passage. The existence of abnormal and transient ionisation of the upper atmosphere in the path of a meteor is, indeed, directly demonstrated by the new radar technique of observing these bodies, whose introduction dates from World War II.

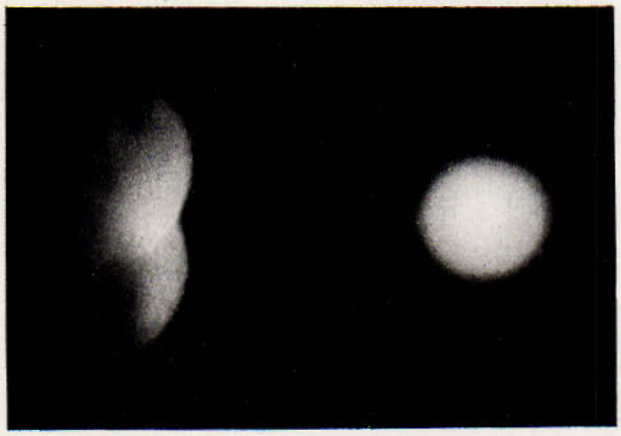


Fig. 475 — NEBULOUS METEORS, AFTER F. BALDET

The principle of this method is analogous to that employed in making soundings of the ionosphere, already described in the chapter dealing with the Earth — the reflection of the radio waves from the transmitter by an ionised layer in the upper atmosphere and their return to the point of origin, where they are recorded as an echo whose strength is an index of the degree of ionisation in the reflecting layer. Thus in the case of meteors it is not quite true to say that the meteor is detected by radar: it is rather the trail of ionised gases in its wake — whose ionisation persists for some little time — that reflects the waves.

Information of the greatest importance concerning meteors and their trains has in recent years been obtained by this method, which in scope and accuracy surpasses the classical visual and photographic techniques. It now appears to be beyond doubt that the light emitted by persistent meteor trains is the result of ionisation, probably also associated with the progressive recombination of the gaseous ions within the train.

It may be appropriate at this point to mention another class of phenomenon — the so-called 'nebulous meteors'. These are ill-defined objects which, like ordinary meteors, appear suddenly and may suffer considerable modifications during their passage across the sky; their paths may be long or short, their angular velocity is frequently low, and their apparent sizes are (comparatively speaking) enormous — surpassing, on occasion, that of the Moon. They are usually very faint: so much so, indeed, that if one happens to pass in front of the brighter regions of the Milky Way, it appears as a dark, cloud-like object.

These curious meteors — first studied systematically by the French astronomer F. Baldet — are of extremely rare occurrence, and the paucity of the observational records makes it difficult to form any definite conclusions about them. The first problem to be solved is their relationship with ordinary meteors.

In short, these phenomena still present us with many unsolved problems even though they occur in the Earth's immediate vicinity. Fortunately, we are better informed about the physical bodies themselves, which produce the

observable phenomena, and also about the conditions under which they encounter the Earth.

The number of meteors. — The casual observer, even if he has paid no particular attention to the matter, is aware that meteors are tolerably frequent. A careful observer will, in fact, see an average of 4 to 10 per hour; this rate will be repeated night after night, and when we consider the whole of the Earth's surface it becomes apparent that their total number must be truly prodigious.

The figures obtained are by no means precise, since necessarily they can only be provided by observers in the occupied parts of the world and under favourable meteorological conditions. Nevertheless, extrapolation from the available observational data permits a sufficiently exact calculation of at least the order of magnitude of the total number that become enmeshed in our atmosphere daily.

It must be pointed out, to begin with, that the easily seen meteors which attract the attention of a casual observer represent only a tiny minority. Across the sky streak an unbelievable number of meteors too faint to be seen with the naked eye: their observation requires telescopes or binoculars of wide field and considerable lightgrasp. Thus two categories of meteor are to be distinguished: naked-eye, and telescopic.

Naked-eye meteors: the category whose numbers are the more accurately determined. The total number that is visible from the whole of the Earth's surface runs into something like 10^8 per day, of which about 500,000 are as bright as the brightest stars.

Telescopic meteors: their numbers increase with decreasing brightness to the limit at which they can at present be observed. The data are not precise since work in this field has only been undertaken comparatively recently, but they show that the daily number of these meteors must be of the order of several times 10^9.

As for the total mass represented by this immense number of individual meteor particles, it was at one time held that it could not amount to more than some 4 — 5 tons daily; we now know, however, that this estimate was much too low, and the best modern investigations suggest that a more precise total is of the order of a hundred tons a day.

Relative frequency and apparitions of meteors. — Although meteors can be seen flashing across the sky at all times, their frequencies are not regular. Generally speaking, they are much more numerous during the later hours of the night, besides which they are particularly abundant at certain times of the year, repeated annually, when the Earth encounters 'showers' or 'swarms' which appear to arrive from well defined quarters in space. Let us examine each case in more detail.

The increase in the number of meteors visible towards the end of the night is a result of the Earth's orbital motion. In order to explain the principle involved, let us suppose that meteorites are scattered uniformly throughout space, that their directions are random, and that their velocities are more or less equal. Through this cloud of fragments the Earth drives its way on its journey round the Sun. Then the hemisphere facing in the direction of the Earth's orbital motion will encounter those meteorites moving in the opposite direction or in directions inclined more or less obliquely to its own; the opposite hemisphere, however, will meet only those meteors whose more rapid motions allow them to overtake the

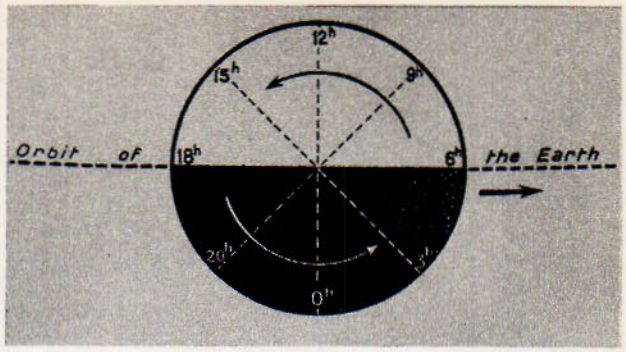


Fig. 476 — METEORS APPEAR IN GREATER NUMBERS TOWARDS THE END OF THE NIGHT, OWING TO THE EARTH'S ORBITAL MOTION

Earth. It can be seen from Fig. 476 that the zone of the Earth's surface over which the night is ending and the day beginning is that most favourably placed for the reception of meteors. (Since, naturally, we can only observe those meteors that arrive during the hours of darkness, the diagram omits those which should be shown on the sunlit portion of the globe.)

The arrival of exceptional numbers of meteors on the same dates each year determined by the particular orbits in which they travel round the Sun. At times, meteors follow one another so rapidly that they give the impression of a veritable rain of luminous streaks — hence the commonly employed term, 'meteor shower'. Some showers in the past have been so rich that the meteors seemed to their observers to be as numerous as snowflakes in a blizzard. In general, the members of such a shower appear (whether they occur simultaneously or in succession) to emanate from the same point on the star sphere; their paths diverge in all directions from this point, called the radiant, and the nearer they are to it the shorter they are. The reason for this is easy to understand. Fig. 477 shows a number of bodies travelling on parallel courses, directed towards the observer O. Owing to the laws of perspective it would appear to this observer that the paths diverge from a point in space which marks the direction from which the bodies are coming; in the same way, the telegraph poles, walls, fences, etc., lining a long straight road appear to converge upon a single point on the horizon.

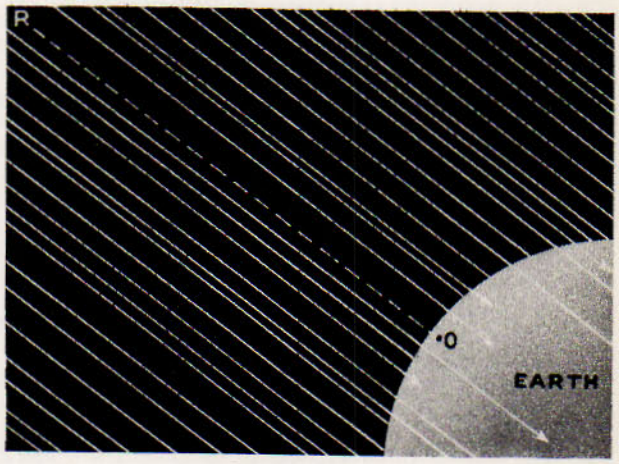


Fig. 477 — The paths of the meteors being parallel, R will appear to an observer at O as the radiant from which they diverge, owing to the effect of perspective.

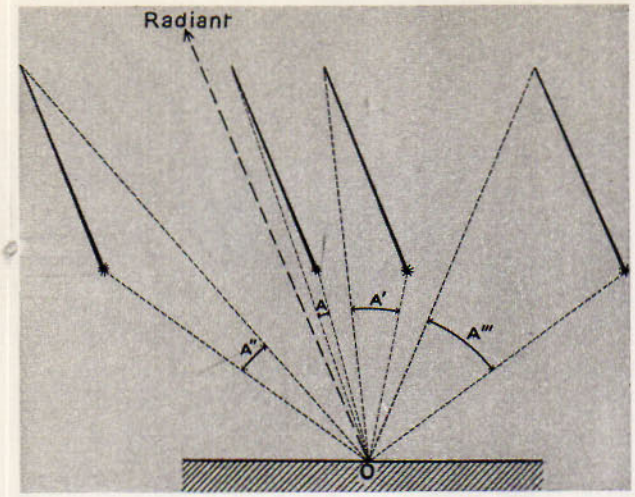


Fig. 478 — It is also an effect of perspective that the meteor paths observed from O appear to be shorter the nearer they are to the radiant.

Meteor systems of this sort are designated by the name of the constellations containing their radiants. Thus the well-known meteors of early August are named the Perseids (radiant in Perseus), those of mid-November the Leonids, etc.

The innumerable particles comprising such a system thus travel through space together, forming, as it were, rivers of dust flowing through space. Since we plunge into each of these streams on the same date each year, this must mark the arrival of the Earth at the intersection point of the two orbits.

Nothing more need be said on the subject of intersecting orbits, since this has already been dealt with in connection with comets and their possible collisions with the Earth. Furthermore, a close relationship exists between comets and periodic meteors, for there is very good reason to believe that the latter are débris from the former.

This connection was first established by the Italian astronomer, Schiaparelli. We have seen that meteors have high velocities; these are generally comparable with the velocities of comets at the same distance from the Sun as the Earth. Thus meteors and comets have, in theory, analogous motions. But more than this, when we determine the orbits along which these swarms are travelling when they encounter the Earth, we find that in many cases they are identical with those of known comets.

Cometary orbits would thus seem to be, in a manner of speaking, materialised in space by cosmic particles scattered all round them; we are logically led to regard these particles as débris from the comet. And, indeed, this relationship between comets and meteors applies as well to extinct comets as to those still active. The following examples demonstrate this.

The most celebrated instance is that of Biela's Comet. We have already seen how this comet, with a period of 6.62 years, was observed in 1846 to have split into two components. It was still in this state at its next return, in 1852. After that, nothing was seen of it at what should have been its returns of 1859 and 1866. It was consequently forgotten until, on 27 November 1872, a most astonishing meteor display occurred. Exactly on that day the Earth crossed the orbit of the vanished comet, which, had it still existed, would have passed this point a few weeks earlier. The conclusion is inescapable that the Earth was passing through the main body of detritus produced by the disintegration of the comet, and this was continuing to circle the Sun in the same orbit and with the same period. A considerable amount of this débris must be distributed round the whole orbit, however, since some meteors are observed every year (and not only at 6-year intervals) on this date. Another particularly rich shower was seen in 1885.

It will be remembered that a similar spectacle was seen on 9 October 1933, some months after the passing of the comet Giacobini-Zinner; it was repeated on the night of 10 October 1946, following another passing of the comet. It may also be pointed out again that the meteors which appear every year on about 6 May are the products of Halley's comet, which since antiquity has appeared punctually every 75 years. Finally, the showers of early August and mid-November are associated, respectively, with Tuttle's comet of 1862 (period, 121 years) and Tempel's of 1866 (period, 33 years), which was seen in that year for the last time.

Although the conditions governing the apparitions of the periodic meteor showers are well understood, it must not be supposed that the displays thus produced on certain fixed dates are always of equal richness. As we have just noted in the case of Biela's Comet, the particles are on the whole distributed extremely unevenly round the orbit: along some stretches the material is scattered very thinly, whereas elsewhere it forms more or less dense condensations. Thus a given meteor swarm will in some years produce a rich display and in others a weak one, and these annual irregularities are quite unpredictable. This state of affairs is easily enough explained: on the one hand, the meteor system will with the passage of time be slowly exhausted by the losses it suffers to the various planets in its path; on the other, even near approaches to planets (without actual encounters) will set up perturbations which will modify to some extent the orbits of its components. Thus the conditions of the swarm's rendezvous with the Earth (including the date) will suffer continual modifications.

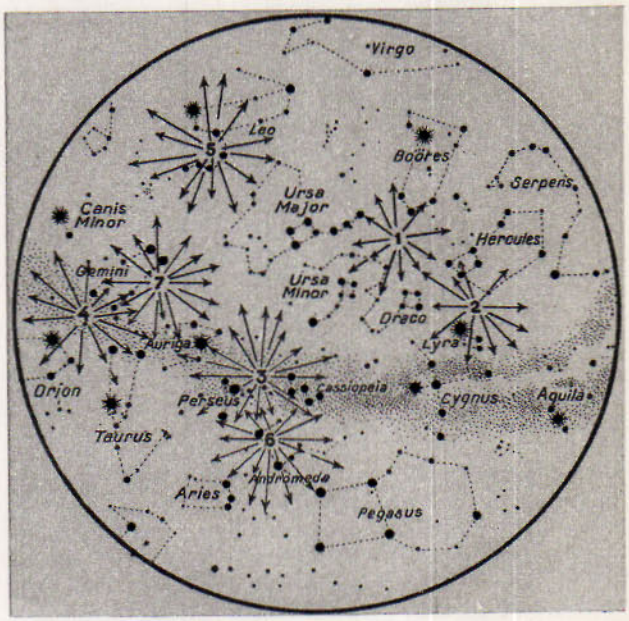


Fig. 479 — POSITIONS OF THE PRINCIPAL RADIANTS OF THE NORTHERN CELESTIAL HEMISPHERE. Their names and dates of appearance are: 1, Boötids (2–3 January); 2, Lyrids (19–22 April); 3, Perseids (9–14 August); 4, Orionids (16–22 October); 5, Leonids (14–18 November); 6, Andromedids (17–23 November); 7, Geminids (9–12 December).

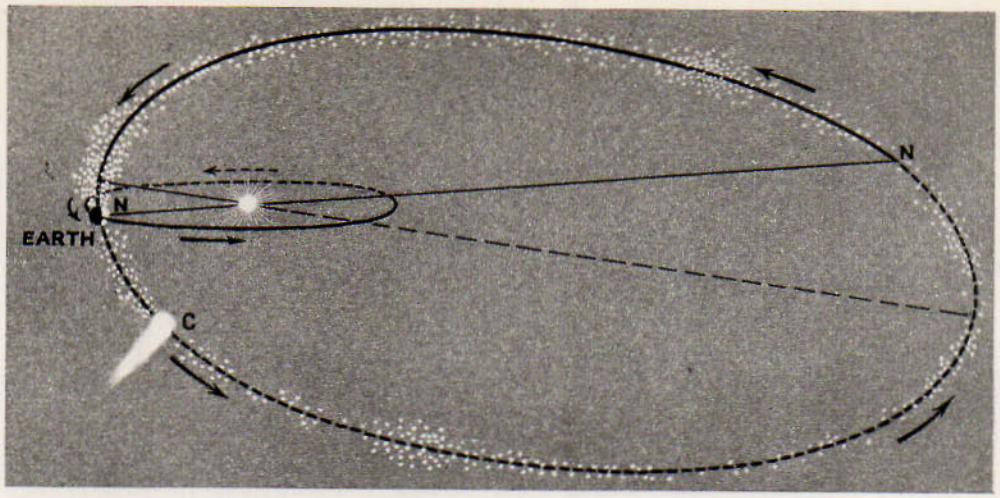


Fig. 480 — METEORS AND THE EARTH MEET AT THE POINT OF INTERSECTION OF THEIR RESPECTIVE ORBITS. The diagram illustrates the meteors associated with Comet Giacobini-Zinner, shown here at position C which it occupied on 9 October 1933, the date on which the Earth arrived at the node N and crossed the meteor stream scattered along the comet's orbit.

If we were to restrict our attention to these facts alone, it would be legitimate to conclude that — as was for a long time believed — the origin of all meteors must be sought in the disintegration of comets. Only the circumstances of their annual apparitions would differ — sometimes we might be plunged into still well defined streams associated with various known cometary orbits, at others we might encounter a sparse collection of widely spaced particles, the remnants of a once rich stream whose association with some long-vanished comet will never be known.

This attractively simple picture has had to be modified, however, in the light of later discoveries, whose implications are of great interest from the viewpoint of cosmogony. The validity of the results already quoted has in no way been undermined: it remains well established that the majority of the meteors visible to the naked eye, and, generally, those associated with the periodic showers, revolve about the Sun in ex-cometary orbits. But we must not overlook the multitude of other, fainter meteors of all sorts, which evade ordinary observation and are only to be seen with the aid of a telescope.

Systematic observations in this field were initiated in 1930 by the Harvard College Observatory expedition to Arizona, whose aim was to undertake a fundamental investigation of this important problem. We must also, in this connection, mention the work of Öpik, Hoffmeister, Muller and Becker, whose main conclusions are incorporated in this account.

It soon became evident that many meteors have velocities far exceeding those possible for bodies moving in elliptical orbits: in other words, their orbits must be either parabolic or hyperbolic. These different types of orbit have already been defined: for a given solar distance, each has its own permissible velocities. Suppose we take as standard the velocity of a body moving in a parabolic orbit at the same mean distance from the Sun as the Earth — 26 m.p.s. Then a lower velocity would indicate an elliptical orbit, a higher velocity a hyperbolic orbit. We have also described the characteristics of these different orbits: a body moving in a ellipse revolves continuously around the Sun; even though it may be difficult in practice to differentiate between an elliptical and parabolic orbit — a parabola being only an ellipse with an infinite major axis — a hyperbolic orbit is very different, since a hyperbola is the figure described by a body falling under the gravitational attraction of the Sun, which it approaches, swings round once, and then makes off again in a different direction, never to return.

Thus, in theory, those meteors with velocities equal to or less than 26 m.p.s. must belong to the Solar System, while those with greater velocities merely pass through it, coming from interstellar space and then receding into it once more.

The fall of meteorites. — Most meteorites entering the Earth's atmosphere are completely vaporised far above the Earth's surface: their arrival at ground level is therefore an extremely rare event — at any rate, so far as meteorites of any size (whose impact on landing has considerable physical and other effects) are concerned. For we shall see that, in addition to such meteoric 'bodies', the detritus of ordinary meteors is all the time falling to Earth in the form of a fine dust, after having floated for some time in the atmosphere.

By reason of the great altitudes at which they appear, meteors are visible over a very wide territory. The arrival at ground level of a meteorite, on the other hand, is necessarily a local event, and one which, if it is not to escape notice, must occur in a region where it will leave a recognisable trace; from time to time, however, a meteorite is by good fortune recovered. Many impacts, since they occur in the wastes of the ocean or in unpopulated regions, must necessarily remain unknown.

For this reason the ascertained data concerning the number and size of meteorites reaching the Earth's surface must be far short of the true figures; nevertheless they can be used to fix our ideas to some extent. Let us see, therefore, what has been established by the extremely careful researches of astronomers in America. It must be remembered that in most cases these bodies have been partially vaporised before they reach the Earth's surface; in others, they are more or less fragmented by explosion, or even reduced to finely divided débris as a result of almost complete disintegration.

It must also be noted in this connection that meteorites do not strike the ground with the fantastic velocities at which they were originally travelling. Experimental data regarding the resistance of the air are lacking in the case of such velocities, but one may logically envisage a more or less powerful braking action, depending on the size of the meteorite and the angle of its path through the

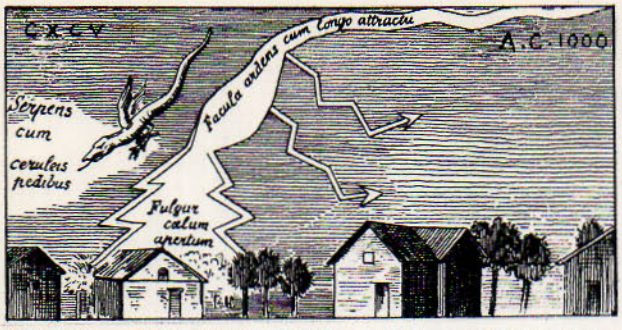


Fig. 481 — THE SUPERSTITIOUS TERRORS OF ANTIQUITY AND THE MIDDLE AGES ARE ILLUSTRATED IN THIS OLD ENGRAVING SHOWING THE FALL OF A FIREBALL IN A.D. 1000

atmosphere; some of these cosmic fragments will therefore finish their journey through the atmosphere in free fall — i.e. under the action of gravity alone, depending on their mass. This is not to say that the shock of impact — equivalent to that of any heavy object falling from a great height — may not on occasion be highly destructive.

These preliminary points having been made, let us return to the actual results furnished by the statistics in question. First we shall consider those meteorites which must have weighed more than 10 lb. before their entry into the atmosphere — their mass being deduced from their brightness as meteors. The number of these falling on the whole of the United States annually has been estimated as about 90, of which 10 or so are observable from populated centres. Generally it is their fragments, weighing a few ounces, which reach ground level.

Now let us consider the case of really large meteorites, weighing, for example, 5 tons or more in space. Of these, about 2 fall every 3 years over the whole of the United States, and 1 every 100 years on an inhabited area; normally the meteoric masses that reach the ground do not much exceed 1,000 lb.

Estimates have also been made of the number of really large meteorites, whose fall to Earth is even more spectacular. The figures for meteorites in this category — estimated as weighing, originally, at least 50 tons (6 to 10 feet in diameter, according to their composition) though reduced to a few tons on arrival — are one every 30 years over the whole of the inhabited globe, or one per 500 years on the United States. Their passage through the atmosphere, like that of the preceding category, is accompanied by violent explosions, their impact causes more or less severe earth tremors, and they bury themselves deep in the ground.

Finally, those extra-terrestrial projectiles whose impact with the Earth can only be described as cataclysmic — masses weighing 200 to 300 tons. These blast enormous craters and spread devastation over a wide area of the surrounding country. It is estimated that the fall of a body of this size occurs, on the average, once every 150 years somewhere on the Earth's surface, once every 2,400 years somewhere in America, and once every 160,000 years on a town or city in the United States.

Still larger meteorites, of the order of 50,000 tons, strike the Earth only every 100,000 years or so. For the United States, successive falls of this category would be separated by an interval of 1,500,000 years, while the chances of a city being struck are something like once per 100,000,000 years.

These estimates are naturally very approximate, since many meteoric falls, as we have said, escape notice at the time of their occurrence. Such statistics do, however, give some idea of the minimum order of magnitude of the bombardment to which our Earth is subjected all the time. But although the last category of meteorites mentioned would indeed constitute a peril of catastrophic proportions, town dwellers need not be unduly perturbed, seeing their extreme rarity; one could wish that they had nothing worse to worry about.

But the rarity of such an event does not detract from the gravity of the consequences for the surrounding countryside when it does occur. Falls of this sort have in fact occurred in the past, and provide impressive evidence of the havoc that meteorites can at times cause. To such a fall, in the far-distant past (estimated as some tens of thousands of years ago) must be attributed 'Meteor Crater', in the Arizona desert. This is an enormous, roughly circular pit, 1,360 yards in diameter and about 650 feet deep, which appears to have been blasted by a tremendous meteorite — probably a lump of iron and nickel, now buried (probably in fragments) about 400 feet below and to one side of the present floor of this gigantic crater.

Equally formidable was the fall that occurred on 30 June 1908 in Siberia. Although no remains of the meteorite itself have been found, the authenticity of the fall seems to be well established by the varied and widespread effects of the impact. Accounts of what happened paint the following picture. The morning of 30 June was cold and clear in the Krasnoiarsk region, where the meteorite fell. Some thousands of people in the area between the Yenisei, the Lena and the trans-Siberian railway later bore witness to 'the flight of a ball of fire across the sky', followed by loud thunder claps and a stupendous crash. These explosions were heard up to distances of about 600 miles. Nearer than this, at distances of 100 to 150 miles or so, men and horses were blown over, while others were knocked unconscious. Enormous waves formed on the rivers, and houses and walls collapsed. From a distance of 250 miles from the point of impact a jet of flame was seen to spurt up into the sky, reaching a height which the eyewitness descriptions suggest must have been about 12 miles. Earthquake shocks were recorded, the waves travelling several times round the world before being finally damped out; similar atmospheric waves were registered by barographs. On the following night, from the Yenisei to the Black Sea, and even as far away as the Atlantic, the sky was abnormally bright, owing to the presence of beautiful silvery clouds at an altitude which must have been in the region of 50 miles.

Situated in a wild and still unexplored region, the scene of this catastrophe was not visited until 1927, when an

Fig. 482 — MASS OF IRON WHICH FELL AT BENDIGO, BRAZIL. Weight, about 14,000 lb.

Fig. 483 — THE METEOR STORM OF 9 OCTOBER 1933. All the meteors appear to diverge from the same point — the so-called 'radiant'.

Fig. 484 — THE ARIZONA METEOR CRATER. *(Photo, American Museum of Natural History)*

Fig. 485 — THE SIBERIAN FOREST DEVASTATED BY THE BLAST FROM THE METEORITE OF 30 JUNE 1908

expedition was organised by the Russian scientist L. Kulik. Aircraft flew over the area, and although they could not land they brought back reports of incredible devastation, already partially camouflaged by new vegetation. Over an area at least 60 miles in diameter the forest was utterly destroyed, the trees flattened and charred, their trunks radiating outward from the point of impact; at the centre of the area there were seen to be more than 200 craters, caused by the detonation of this fantastic fireball.

From a consideration of all the available facts, Professor Kulik estimated that the total mass of the meteorite could not have been less than 40,000 tons. Prolonged and detailed search with the aid of modern prospecting techniques could no doubt reveal what exactly has become of it. But the undertaking would be difficult since the region is practically inaccessible except by air and there are no suitable landing places in the region, while the nearest practicable ports are too distant.

Determinations of the conditions of its encounter with the Earth indicate that it came from the direction in space marked by the constellation Cetus. This gigantic meteorite, probably moving in a hyperbolic orbit, must have crashed into the Earth with a velocity greater than 30 m.p.s., after having travelled for some 300 miles through the atmosphere. It has been pointed out that, had it arrived 4h 47m earlier it would (by virtue of the Earth's rotation) have landed on Leningrad (then St Petersburg) — with consequences which in pre-atomic days would have been without precedent.

Similar meteoric falls, though on a smaller scale, have occurred more recently. It was reported from Nairobi, in East Africa, that at the end of 1946 — the exact date is not known — an enormous explosion in the semidesert region north of Mount Kenya had flattened villages over an area 60 miles in diameter. This was attributed to a meteorite; in any case, one village was set on fire by a hail of meteorites and all the cattle were destroyed, though the villagers themselves managed to escape. At the beginning of 1947 a meteorite weighing several thousands of tons again fell in eastern Siberia.

Finally, the largest meteoric mass so far discovered — in the Adrar district of French West Africa — is estimated to weigh a million tons: it may almost be regarded as a small cometary nucleus.

Apart from these exceptional meteorites, whose impacts are capable of producing real cataclysms, there are the ordinary 'small calibre' meteorites which produce no dire results and which are extremely valuable in providing samples for examination and analysis. These fragments are the only heavenly bodies, apart from the Earth, that we have so far been able actually to touch or handle.

Meteorites may fall in one piece or in several, as a result of explosions whose frequency we have already noted. In the latter case the number of fragments varies widely, from two or three to several score, hundreds, or even more, when a considerable territory may be covered by this celestial shrapnel. As with shrapnel, too, the affected area is elliptical in shape, its major axis lying in the direction of the meteorite's motion; a sieving effect also comes into play, the more massive fragments falling first.

One of the more celebrated showers of this sort is that which occurred at Laigle, in France, on 26 August 1803, when about 3,000 fragments were scattered over an area 6 miles long by 2 ½ wide. Even more spectacular was the Pultusk (Poland) shower of 30 January 1868, when something like 100,000 fragments fell in an area even smaller than that at Laigle.

Although in general it is true to say the arrival of these celestial projectiles cannot help affecting the region involved more or less seriously, it is nevertheless a fact that the consequences are often surprisingly light, owing to

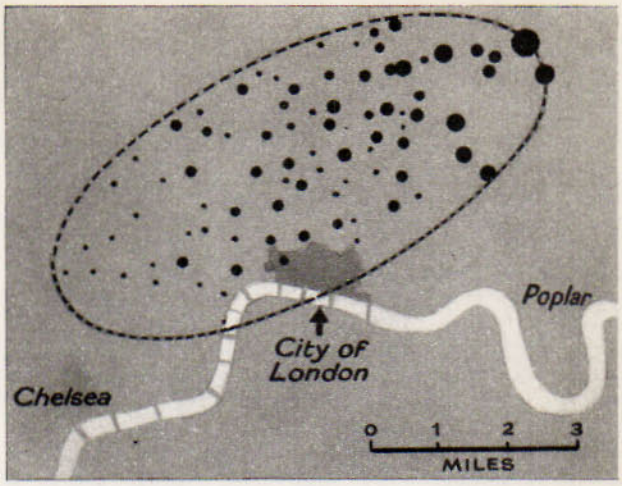


Fig. 486 — AREA OVER WHICH THE 'HAILSTORM' OF METEORITES FELL AT LAIGLE (FRANCE), COMPARED WITH THAT OF GREATER LONDON

the deceleration that the meteorite has undergone while passing through the lower levels of the atmosphere.

Not only are the impacts of meteorites often without catastrophic effects, they are also extremely rare, considering the number of meteors that are all the time streaking across the sky. But although the vast majority of these are completely vaporised in the upper atmosphere, they none the less leave a certain amount of débris behind them; this, in the form of a very fine dust, ultimately sinks to the ground. There is plenty of evidence to show that this fallout of microscopic meteoric particles is going on all the time. It is not perceptible to casual observation, as is the fall of the more spectacular meteorites, since only the most detailed microscopic investigations can yield information about these particles: for though they are raining down upon us incessantly they are invisible to the naked eye.

Owing to their smallness, they remain suspended for a long time in the atmosphere. There they mingle with other particles of every sort (mineral and organic) which are carried aloft from ground level by atmospheric convection. Precipitation (rain and snow), in particular, collects all these particles together, later to release them again after evaporation.

It would thus appear difficult to differentiate between

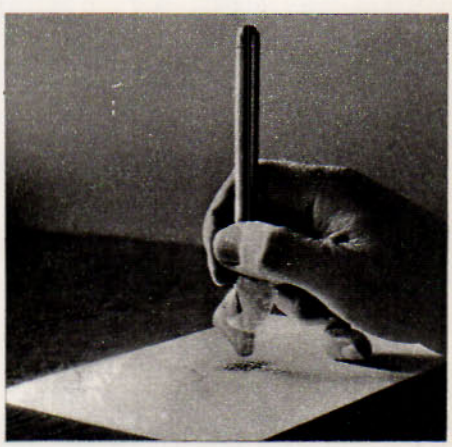


Fig. 487 — SEPARATING METAL PARTICLES FROM THE COLLECTED DUST BY MEANS OF A MAGNET

the cosmic detritus and the particles of terrestrial origin, natural or industrial, which abound in the inhabited parts of the world and which, moreover, are carried far and wide by large-scale atmospheric movements. However, the collection of particles falling directly from the atmosphere during dry periods, and of the slight sedimentary deposits of rain or snow, permits us to recognise some at least of the elements that originated in outer space — those that can be separated from the others.

These are ferrous particles, which for the most part have a characteristic appearance, resembling tiny cinders, globules or droplets — in short, fragments that have been subjected to high temperatures. But we cannot be absolutely sure of their provenance from their appearance alone, since much industrial débris is very similar. Their extra-terrestrial origin is chiefly established by the following facts: first these particles are to be found in every region of the globe, whether on the polar ice, in the snow at the summits of high mountains, or in the mud and ooze of the ocean beds. More recently, the sediment that accumulates in rain and snow gauges has yielded up numerous samples. It is particularly noteworthy that this worldwide fallout is not haphazard, but occurs under fairly clearly defined conditions: thut it is markedly heavier at (or, rather, shortly after) epochs when meteors are most abundant; the time-lag is due to the fact that the particles sink only slowly through the atmosphere, at the mercy of meteorological vicissitudes. Such, at any rate, are the conclusions drawn from the systematic investigation carried out over a number of years at Donville Observatory.

The method by which these facts are established is simple. It consists in collecting — once a week, say — the sediment deposited by the air in some suitable container, and then in separating out the ferrous particles with a magnet. Microscopic examination allows them to be counted, and their shape and size to be studied; with very rare exceptions their diameters are between about 0.005 and 0.0005 in.

From the quantitative angle alone, very significant results have already been obtained. Though meteoric dust is falling all the time, its quantity increases after the comparatively rich meteor showers which occur each year at the same times. Furthermore, this general statement has been reinforced by certain specific observations: thus after the passage of a bright fireball an abnormal number of particles will be collected, some of which (comparatively very large) may measure as much as 0.05 in. in diameter. Again, the spectacular meteor display of 9 October 1933 was followed by a rich fallout of dust which continued for several days, the particles being deposited in order of decreasing weight.

Additional information has been gained from simultaneous observations made at widely separated stations. For example, identical results were obtained from samples of meteoric dust collected on the same dates at Donville Observatory in northern France, and by a voluntary investigator in the south of Morocco: the same varying proportion of the meteoric particles, similarity of the predominant forms, same relative dimensions of the different samples.

But if all these observations furnish incontestable proof of the continuous fallout of cosmic detritus, it cannot equally be asserted that all the collected ferrous material has an extra-terrestrial origin, for, as we have seen, it is mixed with other particles produced by industrial plants which are widely dispersed through the atmosphere and which resemble it very closely. To differentiate between them is not easy. But the quantitative variations, related to the visible meteor displays, show that at any rate the majority of the collected particles do in fact originate in space.

It is clear, however, that these facts do not tell the whole story. Only the ferrous components of the atmospheric dust can be separated out by a magnet; and meteorites, as we shall see, are very far from being wholly ferrous. Thus a considerable amount of meteoric débris, indistinguishable from ordinary terrestrial mineral dust, must escape experimental control.

Finally, the attempt has been made to evaluate the total mass of this incessant bombardment by thousands of millions of meteorites of all sizes. These estimates are based on their observed numbers and their size, as derived from their brightness and from what remains after their more or less complete destruction in the atmosphere. Such results can only be approximate, but they give at least some idea of the order of magnitude involved. According to C. C. Wylie, the average annual deposit amounts to 4×10^6 lb. This, referring to the whole of the Earth's surface, is equivalent to a mass of less than 4 ounces per square mile, or — assuming the fallout were allowed to accumulate undisturbed — a thickness not exceeding about an inch in 5,000 million years.

Thus although one may possibly be apprehensive of the serious local effects caused by a meteorite, one can hardly fear being buried alive by this perpetual and virtually impalpable rain of dust from the sky.

Composition of meteorites. — We have seen that meteorites are lumps or fragments of mineral and metallic composition, and that their dimensions on arrival at ground level vary within wide limits, from grains of microscopic dust to very considerable masses. Their composition is no less varied, and provides the basis for their classification into four main types:

1. *Siderites*, whose average composition is 92% iron, 7% nickel, and often a trace of cobalt.
2. *Stony meteorites*, composed mainly of silica and magnesium oxides.
3. *Siderolites*, containing mineral and metallic elements in varied proportions, the latter being granular.
4. *Asideres*, a class represented by only a few specimens. They exhibit certain very special characteristics, including the complete absence of iron. The most remarkable are the 'charcoal' meteorites which fell at Alès, in the south of France, in 1806, and at Orgueil in 1864. Generally speaking, they are lumps of very friable material which break up rapidly under exposure to the weather: this may explain the rarity of the occasions on which they have been found after a meteorite fall.

The stony meteorites are much the most numerous, representing about 95% of all the samples collected, and their examination has uncovered some interesting and significant facts. Their mineral elements comprise a wide range of rocks, many of which are similar to those found on Earth; meteoric metals, however, always have a granular structure, called 'chondritic', which is entirely peculiar to meteors. No new or unknown chemical element, however, has ever been discovered in a meteorite. Taking them all together, the most frequently occurring elements are iron (72%), oxygen (10%), nickel (6%), silicon (5%) and magnesium (4%), with others amounting to less than 1% each.

In addition, meteorites contain various gases which are liberated under the action of heat; these show a certain similarity to the gases of which comets are composed — notably hydrocarbons, carbon monoxide and nitrogen. Hydrogen, also, is abundant in siderites.

Origin of meteorites. — 'Stones falling from the sky' were once rated as prodigies and marvels. Later, they passed from the supernatural realm to that of atmospheric phenomena. Lavoisier, for example, in a report to the

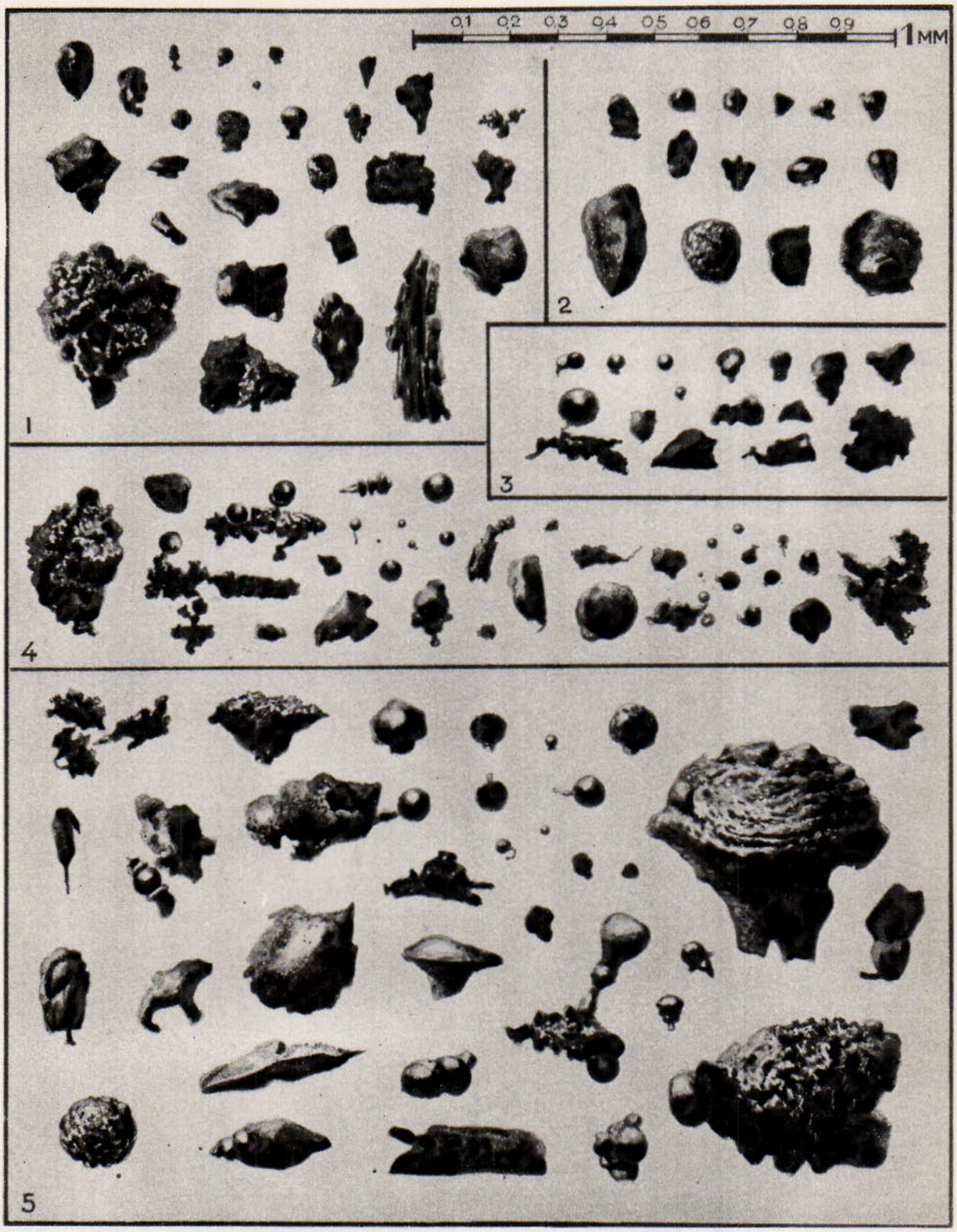


Fig. 488 — SIMILARITY BETWEEN PARTICLES OF MAGNETIC DUST (HIGHLY MAGNIFIED) COLLECTED FROM WIDELY SEPARATED LOCALITIES: 1, 2, from mountain snow (Pyrenees); 3, from the tower at Montlhéry; 4, from a gutter in Paris; 5, from a reservoir at Donville (Manche, France).

Academy of Sciences in 1792, stated that the samples that had been collected were ordinary stones, modified only by the action of heat. It is to Chladni, a physicist of the same period, that we owe the recognition of the cosmic origin of shooting stars and meteorites. But it was only as the result of the rain of meteoric fragments that fell at Laigle in 1803 (which were examined by a scientific mission headed by Biot) that serious attention began to be given to these visitors from space.

It is probable that not all meteorites have the same origin. In the absence of other proofs this would seem to be evident merely from their appearance and the circumstances of their arrival at the Earth's surface. Indeed, certain distinctions seem to have been firmly established between ordinary meteors, single or in showers, and the more impressive fireballs (whether or not these occasion the fall to Earth of sometimes enormous lumps), which occur with no regularity whatever.

In support of this contention, it is difficult to invoke the diversity of composition of the fragments that have been recovered and analysed, since in the first place none of these is unique. However, this has not proved an obstacle to the formulation of various hypotheses, regarding which something must now be said.

It has been suggested that meteorites are the fragments of a world — a planet or an erstwhile satellite of the Earth — which has since been destroyed. Such a body would be composed of many different minerals and metals, to judge from the constitution of our own planet. The meteorites formed by a cataclysmic disruption of this sort would therefore not all have a similar composition since they would derive from different parts of the surface and interior of the parent body.

Similarly, a great diversity of composition is encountered in volcanic material; whence it has been suggested that the origin of meteorites is to be sought in gigantic volcanoes that may once have existed on the Earth, Moon, or other planets.

Hypotheses of this kind suffer from serious defects. The first must not only account for the disruption of the parent world, but also explain the formation of fragments as small as meteors, and their dispersal in the orbits of every conceivable orientation that they are known to be following when they encounter the Earth. The second theory is subject to the same objection, and to the additional one that eruptions of quite stupendous violence would be required to liberate the ejected material from the gravitational attraction of the parent world. Thus, in the case of the Earth, it would have to be expelled with a velocity not less than 7 m.p.s., which is more than 10 times the velocities actually observed during volcanic eruptions.

Furthermore, such theories are irreconcilable with many facts of observation. We have already mentioned the close connection between the periodic meteor swarms and comets, which naturally led to the suggestion that the former are the débris of the latter — since comets (at any rate, those with which we are familiar) seem to have limited life spans. An actual disintegration of the comet is envisaged, its fragments being scattered round the whole orbit. In this way appeal is made not to some unique catastrophe like the disruption of a planet, but rather to a progressive dispersion of the cometary material under the influence of the Sun and of planetary perturbations; recent investigations have in fact suggested that the nucleus of a comet is not a single body but a collection of minute scattered elements, perhaps separated from one another by very considerable distances.

And so, although the origin of meteors is still much of a mystery, it seems to be tolerably certain that we are dealing with objects whose formation occurred in the remote past. As we shall see later, their helium content suggests an age of about 3×10^9 years, which is comparable with that of the Earth. From the point of view of cosmogony alone, this fact is of the greatest significance.

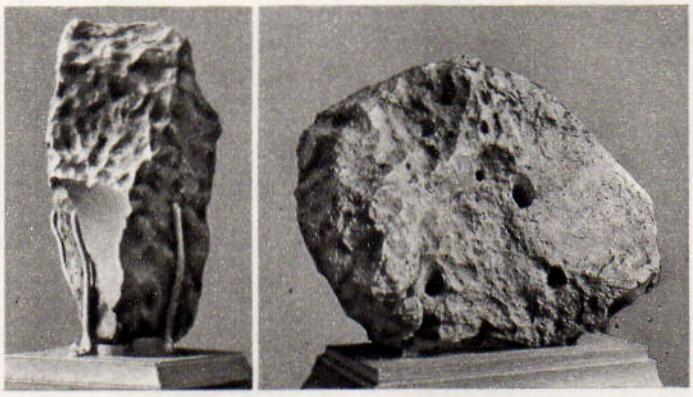


Fig. 489 — LUMPS OF METEORIC IRON

Fig. 490 — STONY METEORITES

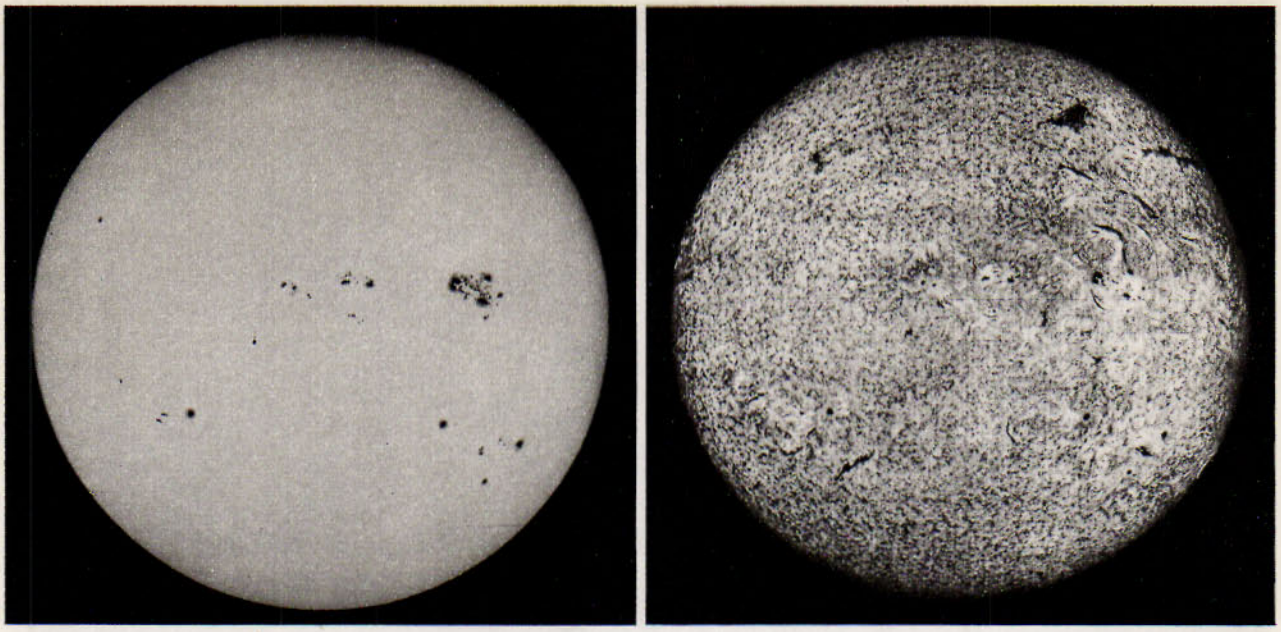


Fig. 491 — SIMULTANEOUS VIEWS OF THE SUN. Left: direct photograph showing the photosphere dotted with sunspots. Right: spectroheliogram showing the upper levels of the chromosphere in the red Hα line of incandescent hydrogen. Note the dark filaments (prominences projected on the disc) and the whirlpool patterns surrounding the spots (solar vortices). *(Photos, Mount Wilson Observatory)*

CHAPTER 10

THE SUN

We have now completed our survey of the Sun's retinue of planets and other bodies that revolve about it. The enormous distances involved, the majestic sweep of the planetary motions, and the physical effects experienced by these worlds — all testify to the paramount gravitational control exercised by the Sun and the intensity of the flood of radiation that it emits. The Sun is not only the central body, but also the prime mover of the Solar System.

We must now give some account of this stupendous source of energy, our Sun — in so far as a satisfactory account can be given at the present time.

To the uninstructed eye, the Sun simply appears to be an intolerably bright disc. In any case, it is naturally regarded as the most magnificent of all the heavenly bodies, and the one, moreover, which most interests us personally, since all manifestations of terrestrial life are dependent upon it. But the Sun's splendour, which at first sight is incomparable, is merely relative — due only to its comparative nearness. The Sun, in fact, is just a star: one among the thousands of millions that populate space, which by reason of their inconceivably great distances appear to us as mere points of light. If the effects of unequal distances are allowed for, the real brightness of the Sun and the stars can be compared, and it is then found that our own star occupies a very modest position in the stellar hierarchy. We shall come to appreciate this point better later on, but for the moment let us confine our attention to nearer home.

Both terms, 'Sun' and 'star', refer to bodies that are utterly unlike those with which we have been concerned up to now. Like its distant brothers, the Sun is self-luminous; and we may amplify this statement, as a first approximation, by saying that it is a gaseous body, strongly condensed towards its centre.

When the Sun is examined with the naked eye — as is possible when its excessive brilliance is reduced by a layer of fog or when it is near the horizon — it appears as a uniformly bright disc with a sharply defined edge. Probably because of this, the Sun was for long ages thought to be a perfect and luminous globe whose purity, immaculateness and serenity nothing could ruffle. The reality is very different.

First of all, however, we must give some idea of the dimensions of the body of the Sun itself.

Dimensions, shape and mass of the Sun. — The body of the Sun (i.e. excluding its atmosphere) has a diameter of 865,000 miles — 109.5 times that of the Earth. One may perhaps form a more concrete idea of such a size by reflecting that it is roughly equal to twice the diameter of the Moon's orbit about the Earth. Thus if we were to imagine the Sun as a hollow sphere with the Earth situated at its centre, then the Moon would be only about half way from the Earth to this immense spherical envelope surrounding it.

If, now, we compare the respective volumes of the Sun and the Earth we find that the Sun's is 1,306,000 times greater than the Earth's. To be precise, its volume is equal to 337,000 million million cubic miles.

The enormous size of the Sun explains how it is that, despite its distance, it still presents a disc of considerable

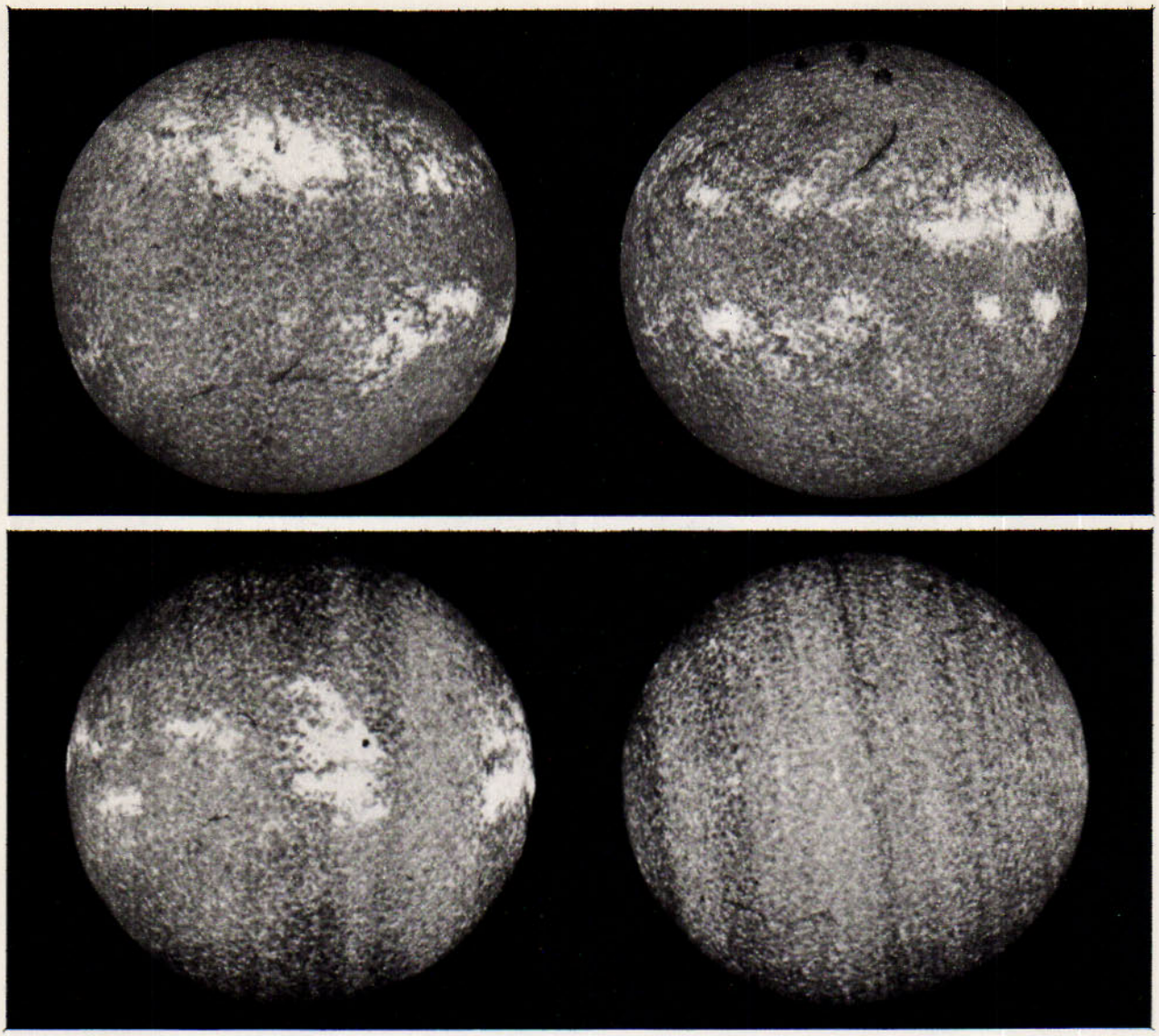


Fig. 492 — SPECTROHELIOGRAMS SHOWING FLOCCULI AND FACULAR PATCHES OF IONISED CALCIUM IN THE UPPER LEVELS OF THE CHROMOSPHERE. Upper left: 21 February 1926, faculae in high latitudes during the increasing phase of the cycle of solar activity; right: 30 May 1928, faculae in mid-latitude at about sunspot maximum. Lower left: 17 December 1929, faculae in low latitudes during the decreasing phase of the solar cycle; right: 21 February 1932, absence of faculae at sunspot minimum. *(Photos, Meudon Observatory)*

angular diameter — 32′ 2″, or rather more than ½° on the star sphere. This, however, is only the mean value, for the Sun's angular diameter varies continuously owing to the changing distance of the Earth as it moves round its elliptical orbit. The limiting values are 32′ 36″ and 31′ 32″, the former being attained at the beginning of January and the latter at the beginning of July.

Unless they are to mislead us, however, these figures must not be taken as rigorously accurate. Measurements of the solar disc are extremely difficult to make with accuracy; furthermore, they are entangled with irradiation effects which tend to falsify the determinations by exaggerating them. Finally, account has to be taken of the vicissitudes of our own atmosphere, for the observations have necessarily to be made through a layer of air whose temperature gradients affect the clarity of the image in different ways. For these reasons, the diameter of the Sun's disc is not known with an accuracy more than about 0.1 or 0.2 seconds of arc. To express this somewhat more graphically, our best determination of the Sun's diameter is uncertain by about 100 miles; nevertheless, this is an almost negligible distance compared with the diameter itself.

Apart from this, periodic variations of the solar diameter — in step with the sunspot cycle, which will be discussed later — have been suspected; according to some observers, they amount to about 0.5″. It has also been thought that in place of the polar flattening exhibited by some other members of the Solar System, the Sun may be slightly elongated in this direction, its polar diameter being some 0.3″ longer than its equatorial. The lack of absolute precision in the measurements — whose magnitude is of precisely the same order as these suspected anomalies — at present prohibits us from giving a definite answer to these questions. It cannot be said that, in the present state of our knowledge, any variations in the Sun's size, or departure from sphericity, have really been established.

Despite its colossal size, the Sun is a comparatively light body. If it were homogeneous (which, as we shall

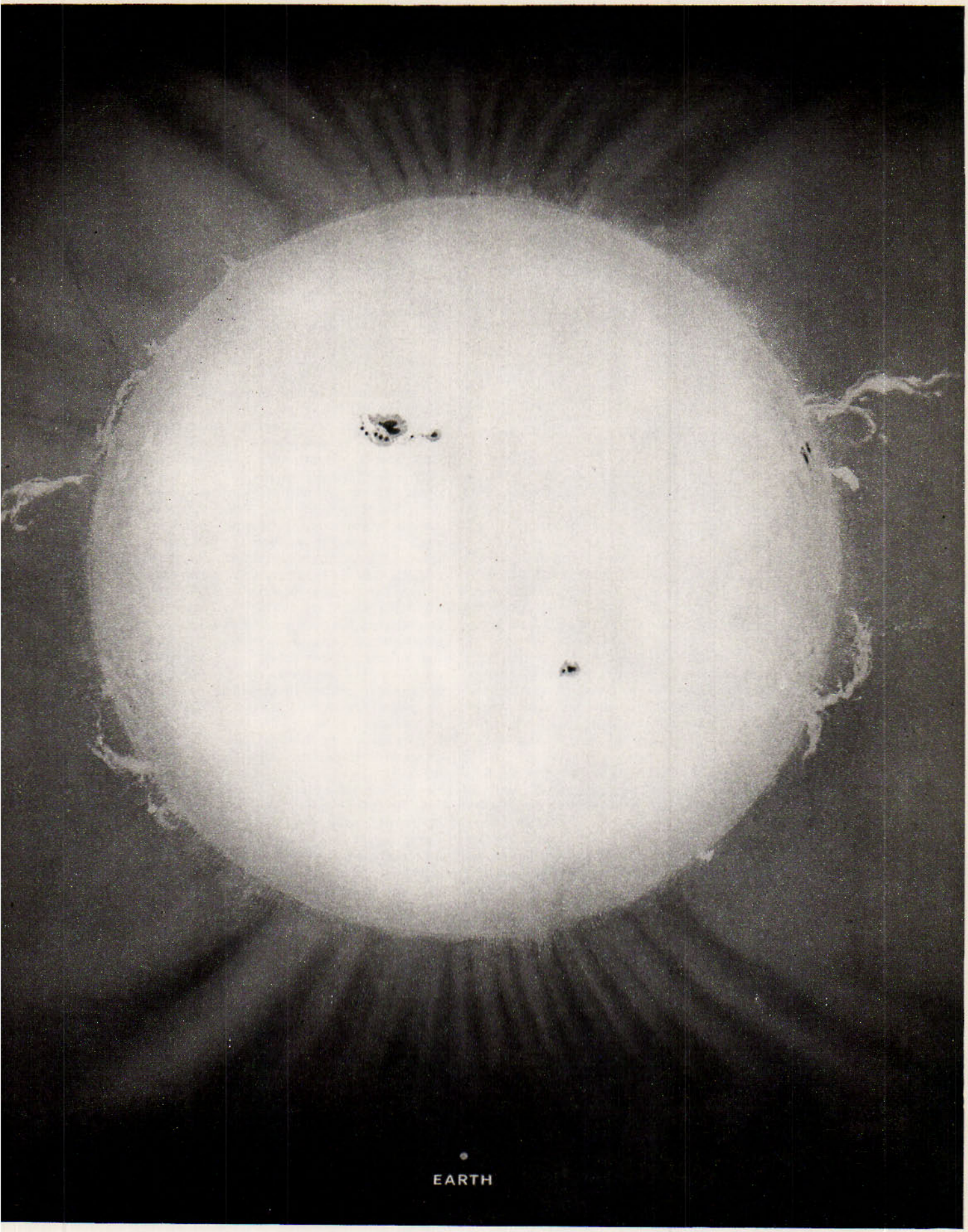


Fig. 493 — THE SUN. The small circle at the bottom of the picture shows the comparative size of the Earth.

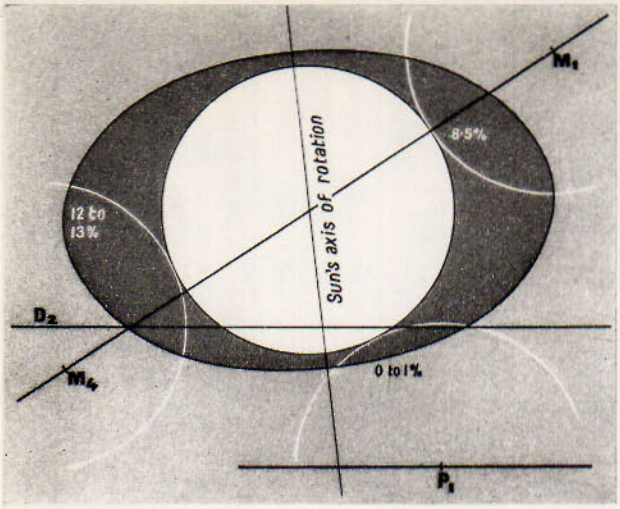


Fig. 494 — GENERAL APPEARANCE OF THE RADIO SUN ON 158 Mc/s, with percentage fading at the optical contacts of eclipses. D, Dakar; M, Makala; P, Paris. (From *Nature*)

see later, it is not), its density throughout would be 1.41 — i. e. about 41% greater than that of water. For although the relative volumes of the Sun and the Earth are 1,300.000:1, the ratio of their masses is only 332,000:1.

Nevertheless, the Sun is still an incomparably more impressive body than any of the planets, even the most massive. And it is not difficult to understand the overwhelmingly powerful rôle that it plays in the affairs of the Solar System, or the vast extent over which its influence is paramount.

The Sun's mass and diameter are such that the force of gravity at its surface is 27.89 times that at the surface of the Earth. If it were possible for a man to be subjected to such conditions he would have the unpleasant experience of weighing 2 tons. At the same time, a body in free fall would cover 446 feet in the first second.

Solar radio observations. — The Sun was the first object to be observed by radio techniques. As early as 1942, British observers had detected the influence of solar flares on the performance of radar sets used to detect enemy aircraft and to guide their own towards distant objectives. This to a large extent explains the enormous development of radio and ionospheric observations during the latter years of the war, and the interest that they aroused in military circles in every country.

Attention may be drawn to the following two points by way of introduction:

(1) It follows from Kirchhoff's law (p. 481) that the longer the wavelength of the observed radiation, the higher and more rarefied the solar levels that radiated it.

(2) The higher the temperature of these layers, the longer will be the wavelength of their radiation.

Hence the radiation detected by radio telescopes 'tuned' to the metre waveband are emitted by the outer corona, although synchronous visual observations refer to the photosphere. There is usually a very close connection between the Sun's general level of activity (as indicated by Wolf's sunspot number, for instance) and the mean intensity of its radio emissions — particularly of its decimetre-wave radiation.

But an equally close correlation (close to within a few minutes) can be established between the appearance of a flare and the occurrence of a very sharp and sudden increase in the Sun's long-wave radiation; many very characteristic observations of this sort have been made (in France, by M. Laffineur at Meudon) on the 55-cm. wavelength. The lag of a few minutes between the visual and the radio observation — a lag which increases with wavelength (and therefore with the altitude of the layer emitting the radiation) — almost certainly represents the time taken for the disturbance to be propagated from the chromosphere to the outer corona. This indicates that there is a certain unity between the phenomena of the photosphere, chromosphere and corona — phenomena which have hitherto only been observed independently of each other.

The resolving power of radio telescopes is poor, and this defect increases proportionately with wavelength; at the present time it is of the order of 1° for decimetre waves. It is therefore impossible to investigate the individual sources of the Sun's radiation — however interesting and valuable such a study might be.

The radio astronomer has for this reason to fall back on observations made during total eclipses, and such observations are, as may be imagined, among the most important that are nowadays undertaken on these rare occasions. The radio telescope is directed towards the Sun, and the diminutions of the instrument's electrical output are recorded as the Moon's disc encroaches first upon the corona and then upon the solar disc itself. A sudden drop in the trace will occur whenever the Moon occults a particular source of radiation on the Sun's surface — notably sunspots. In this way a fairly accurate idea of the location of these sources can be gained, even though they are far from being pinpointed.

One particularly valuable result of this work has been the definition of the 'radio Sun', which is markedly different from the Sun as revealed by ordinary visual and photographic methods. The radio Sun is strongly ellipsoidal, as can be seen from Fig. 494, which combines the data obtained at several successive eclipses, including an annular eclipse observed in September 1951 by a French expedition to Makala, Nigeria, on a wavelength of 3.12 cm.

In addition to this method of attacking the problem of locating individual radio sources, interferometric observations have within the last few years provided relatively detailed information about the radio Sun outside eclipse. These have been developed, in particular, at Cambridge, where the apparatus consists of two observational base-lines, oriented respectively north-south and east-west, which incorporate means for varying the separation of the antennae. In this way, not only the shape but also the distribution of the radio Sun's intensity maxima can be deduced.

New advances are being made almost daily in this, one of the most active branches of contemporary astronomy.

The Sun's brightness. — The Sun's disc, as we normally see it in the sky, appears to be uniformly bright. This impression is dispelled by even the most perfunctory telescopic examination, which will reveal that the brightness decreases from the centre towards the limb. This difference is even more marked in photographs, since the limb regions are not only less intense but are also redder, and therefore less actinically effective than the centre. The darkening of the limb cannot — as was once thought — be explained in terms of a gaseous layer overlying the bright visible surface and absorbing the radiation from the latter in proportion to the thickness of the layer that it has to pass through. It is rather due — as will be explained more fully in connection with the constitution of stellar atmospheres (Chapter 15) — to the depth of the layers emitting the radiation that reaches us;

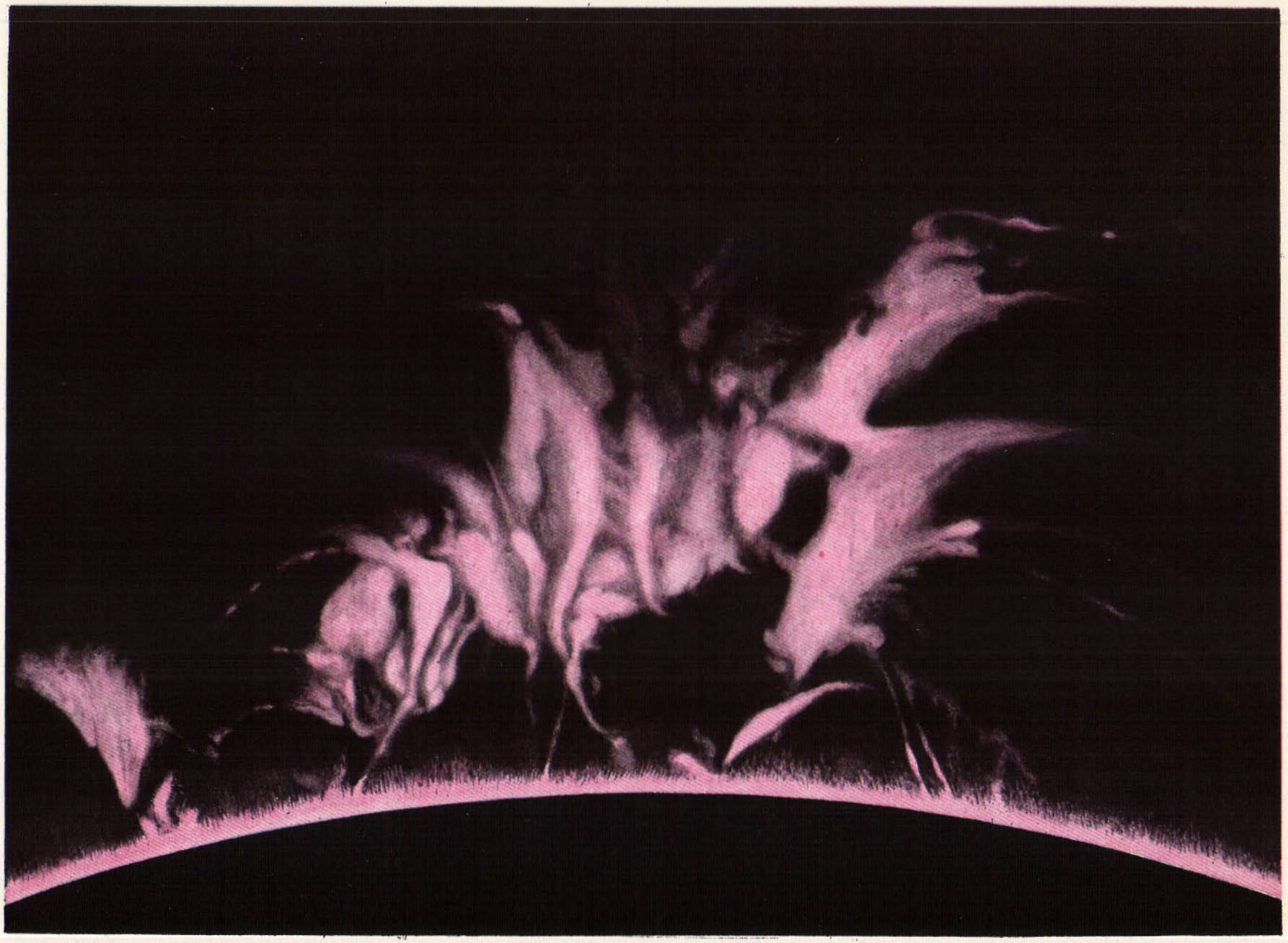


Plate VI. A SOLAR PROMINENCE

this varies from the centre of the disc to the limb in such a way that in the vicinity of the latter the temperature of the effective radiating layer is lower, and therefore the emission weaker, than at the centre of the disc. In any case, it is the brilliant surface — the photosphere — that produces the vastly greater part of the light which we receive on Earth, and without which there would be no daylight; so far does this light outshine even the most brilliant forms of terrestrial lighting that it is difficult to form a true conception of it. The following figures may, however, help towards an appreciation of the Sun's fantastic brilliance.

Many photometric determinations have been made of the light reaching us from the Sun. These are of course concerned with the quantity reaching us at ground level, after having traversed the atmosphere, and are therefore dependent on the Sun's altitude above the horizon at the time; its varying distance from the Earth during the course of the year must also be taken into account. However, the necessary corrections are not difficult to apply. With the Earth at mean distance, and the Sun at the zenith, the illumination received by a horizontal surface at sea level would be of the order of 100,000 lumens according to Fabry, or 103,000 lumens according to Kimball. Employing more familiar terms, this may be translated thus: the measured intensity is equivalent to that of a light source of the same candle-power (i.e. 100,000) situated at a distance of a yard. When the effect of atmospheric absorption is eliminated, the figure increases to 134,000 lumens.

Further calculation, based on this general value, shows that the brightness of the solar surface itself is at least equal to 1,500,000 candle-power per square inch. When one thinks of the number of square inches that make up the surface of a globe the size of the Sun . . . one may, not without reason regard the total as literally inconceivable. Even so, we shall see later that by comparison with many high-temperature stars the Sun is a mere glow-worm, as well as being almost 'cool'.

Temperature and constitution of the Sun. — Let us now examine in greater detail some of the characteristics of the solar furnace.

Chemically, its composition is similar to that of the Earth, but the conditions under which the solar material exists are utterly unlike anything that can be reproduced

Fig. 495 — THE RICE-GRAIN, PHOTOGRAPHED BY JANSSEN AT MEUDON OBSERVATORY. Diameter of solar image, about 35 in.

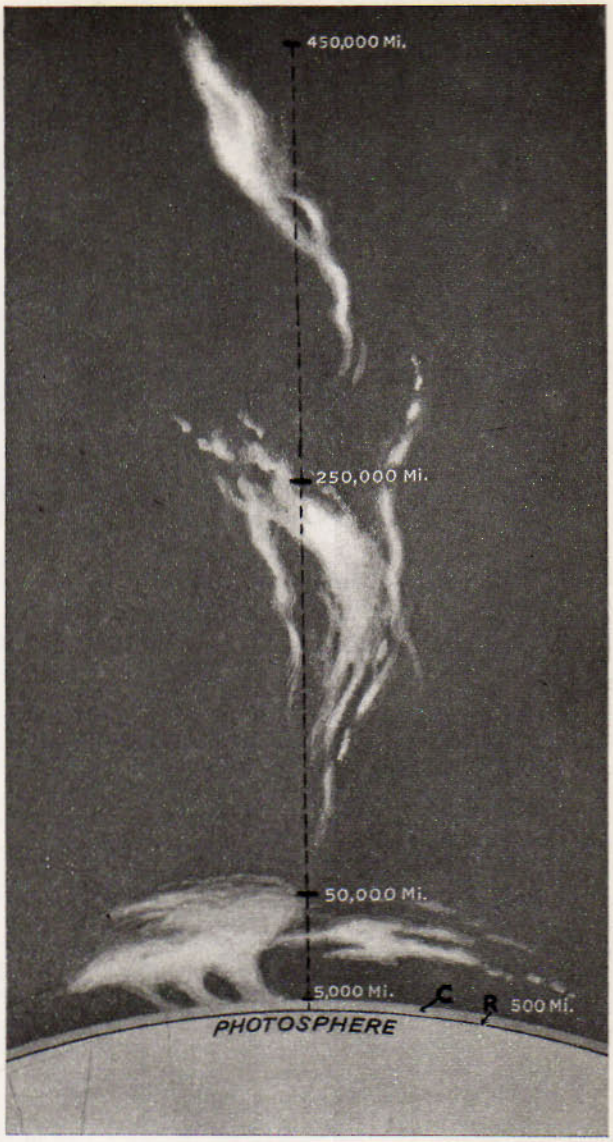


Fig. 496 — DIAGRAMMATIC SECTION OF THE SUN'S ATMOSPHERE: chromosphere, reversing layer and photosphere.

on Earth, even in a hydrogen fusion reaction. The Sun's temperature is about 6,000° C, according to the best determinations made to date; this, at any rate, is the temperature of the visible surface whose radiation we receive on Earth, but it increases steadily towards the Sun's centre, where the calculations of the astrophysicists indicate it must atain a value of several million degrees (see Chapter 15).

To our terrestrial way of thinking, everything should therefore be vaporised. But because of the size and mass of the Sun, enormous pressures come into play, reaching several thousand million atmospheres towards the Sun's centre. The terms 'solid', 'liquid' and 'gas' describe the various states of matter as it is encountered in our terrestrial environment, and our ordinary vocabulary is inadequate for a precise description of the state of the solar matter. In default of an adequate word, some idea of the nature of the Sun's material is best conveyed by describing it as gaseous, since it behaves in accordance with the well-known gas laws of physics. It should nevertheless be remembered that under very great pressures the gases acquire densities and other properties

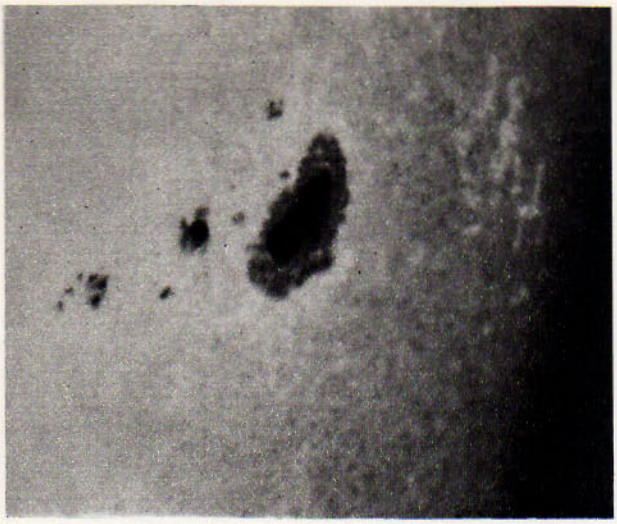


Fig. 497 — SPOTS AND FACULAE NEAR THE LIMB. Diameter of solar image, about 24 in. *(Photos, Meudon Observatory)*

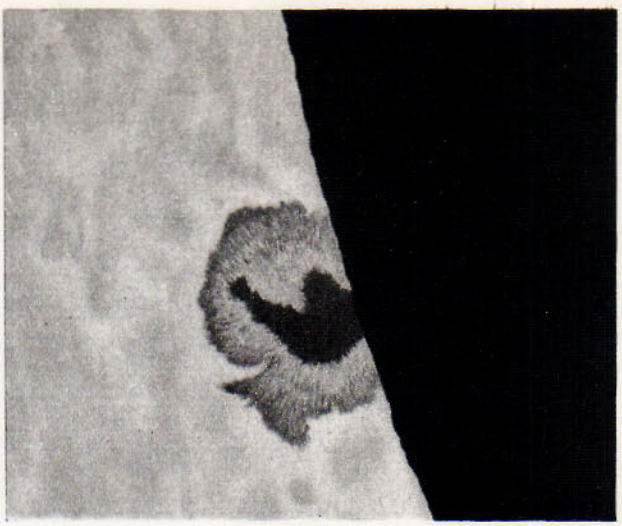


Fig. 498 — THE REAL BRIGHTNESS OF A SUNSPOT IS APPARENT WHEN IT CAN BE COMPARED WITH THE MOON'S LIMB DURING AN ECLIPSE

that make them more appropriately comparable with what we should commonly describe as 'liquid' — or perhaps even 'solid', in the Sun's interior.

However, only the outer layers of the Sun are accessible to direct observation. All our knowledge and suppositions regarding the constitution of its interior are derived from theoretical considerations, and it is generally agreed that at its centre the temperature must be in the region of 20 to 25 million degrees, the density surpassing that of the heaviest metals. (See Chapter 15.)

If practically the whole of the Sun is gaseous, it is necessary to look for an explanation of the sharply defined appearance of its limb. In point of fact, what we see is not a well defined surface, such as a liquid or solid body would have, but a particular zone or level in the solar atmosphere, in which perceptible changes occur. We shall discuss this in greater detail later; here it need only be said that owing to the high value of gravity on the Sun the pressure increases very rapidly the further one penetrates into the gaseous mass. Calculation shows that at a certain level a difference of a score or so miles is enough to double the pressure. It is thus easy to see that the transition from a transparent medium to one of sufficient density to appear opaque would be rapid. In this way the appearance of a bright surface — the photosphere — is produced.

The thickness of the outer gaseous layer is negligible compared with the radius of the solar globe, and by reason of its distance from the terrestrial observer this transitional zone has a practically inappreciable thickness. Thus it is that the more highly condensed 'body' of the Sun appears to have a sharply defined outline. Above this zone, meanwhile, and without any discontinuity, lies the immense gaseous atmosphere, becoming more and more rare with increasing height, which surrounds the Sun.

These outer envelopes of the Sun ordinarily escape notice, for reasons that are not difficult to understand. Because they are transparent, the dazzling surface of the photosphere is readily visible through them; and round the limb, where they should be visible in profile, though much less bright than the disc itself, they are submerged in the greater brightness of the Sun's light diffused in our own atmosphere. Their visibility is (or was, until recent years) restricted to those rare moments when the solar disc is totally eclipsed by the Moon and the difficulties introduced by our atmosphere are temporarily in abeyance. Physics, however, has come to the astronomer's assistance by providing a method for eliminating these atmospheric obstacles, so that at any rate the brightest regions overlying the photosphere can now be recorded, and their phenomena studied, without having to wait for an eclipse. The possibility of continuous observation that

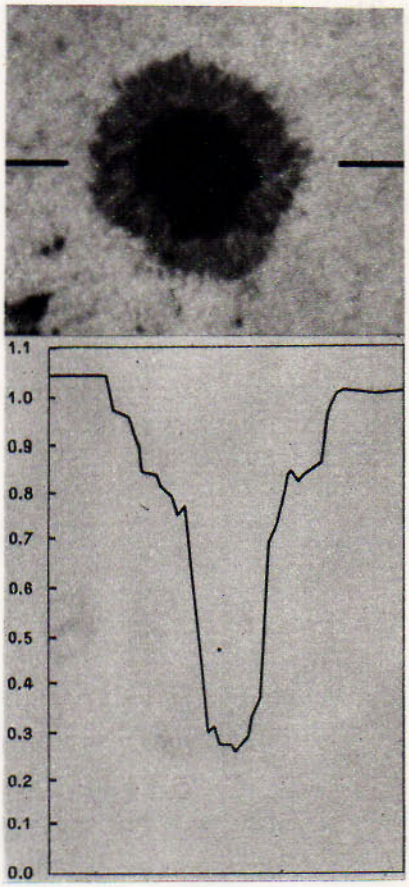


Fig. 499 — VARIATION OF BRIGHTNESS ACROSS A SUNSPOT OF REGULAR TYPE, COMPARED WITH THAT OF THE PHOTOSPHERE

Fig. 500 — LARGE SUNSPOT, VISIBLE TO THE NAKED EYE WHEN THE SUN IS OBSERVED THROUGH FOG

this provides has been responsible for a rapid increase in our general knowledge of the Sun.

The base of the envelope that overlies the Sun's visible surface is a layer called the chromosphere. At the level where this passes into the photosphere a thin stratum known as the reversing layer is located. (The reason for this name will be explained later: Chapter 13.) Finally, above the chromosphere, comes the corona — the outermost envelope, of considerable though varying extent, which is such an impressive spectacle at total eclipses of the Sun.

The photosphere, the chromosphere and the corona are the three main divisions of the observable region of the Sun. Each has its own characteristics, which we shall examine in turn, but it should be emphasised at once that these zones are not independent of one another: the phenomena which are produced in each react more or less intimately on one another, and all are related to the fluctuating tempo of the Sun's activity.

Fig. 501 — LATE STAGE IN THE LIFE OF A SUNSPOT: PHOTOSPHERIC BRIDGES ARE INVADING IT PRIOR TO COVERING IT COMPLETELY

Let us now examine more closely the structure and characteristics of these successive observable layers.

Appearance of the solar surface; photosphere, rice-grain. — Quite apart from the general inequality of brightness from the centre of the disc towards the limb, the surface of the photosphere has a remarkable texture of its own which introduces further differences of brightness. It is, in fact, composed of a multitude of small, bright granules which are superimposed on a relatively darker, greyish background. On account of its appearance, this structure has been called the rice-grain. But the 'grains' are no inconsiderable objects: their diameters range from about 120 to 650 miles. Moreover, they in their turn appear to have a granular structure. Subject to rapid displacements, that bring to mind the seething surface of a liquid at the boil, they have been shown to be ephemeral structures whose individual existences barely exceed a few minutes.

What is the explanation of this characteristic feature of the photosphere? The first idea which suggests itself is that the rice-grain is analogous with the mackerel clouds of our own skies — that is, consists of very hot clouds floating in a relatively dark medium. Nevertheless, their

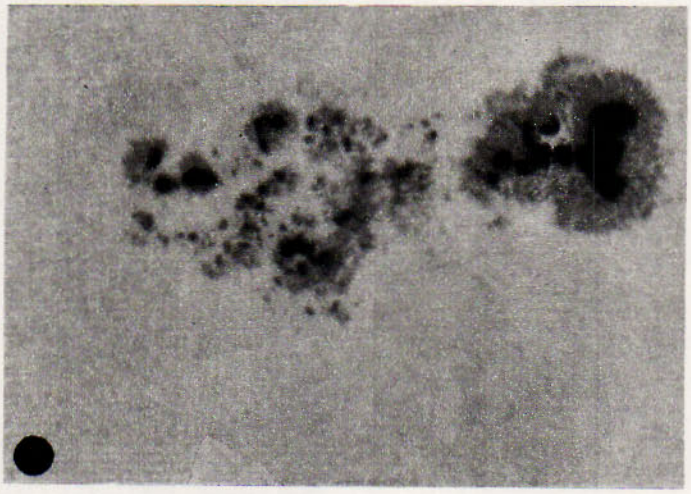


Fig. 502 — A SUNSPOT GROUP, RATHER MORE THAN 100,000 MILES LONG. The black spot represents the Earth on the same scale.

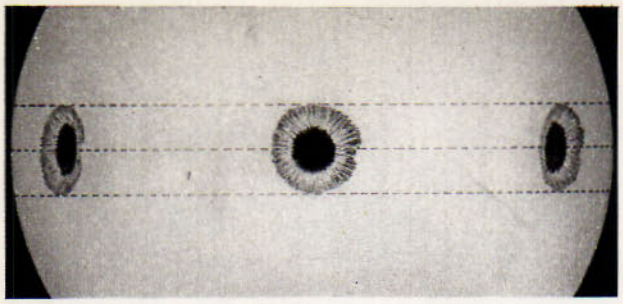


Fig. 503 — DIAGRAM ILLUSTRATING THE APPARENT DISTORTION OF A SPOT SEEN NEAR THE LIMB, OWING TO PERSPECTIVE

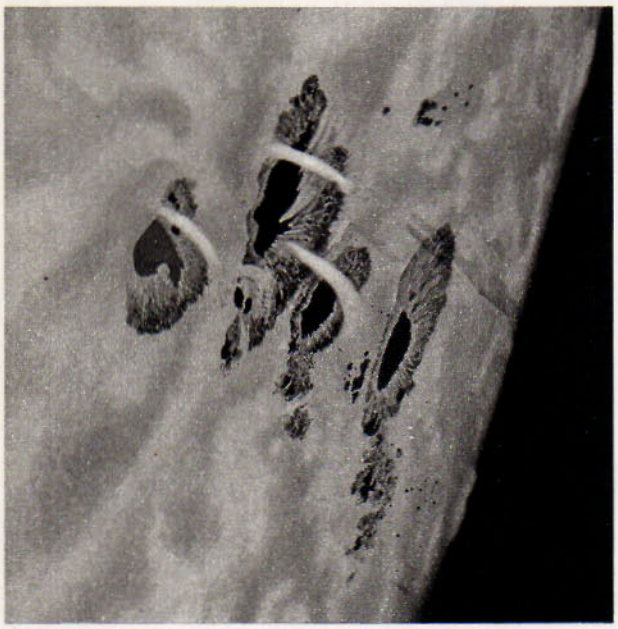


Fig. 504 — FACULAE SEEN IN PROFILE AS PROJECTIONS OVER THE SUN'S LIMB

transistory character and their rapid motions are not to be reconciled with the comparative stability of the terrestrial clouds which they are supposed to resemble. A more plausible idea is that they are the upper extremities of flame-like filaments which break through the lower gaseous levels. The photosphere could in this case be likened to a field of corn; and it is suggested that the ceaseless activity of the mean observed level sets up wave-like undulations whose troughs correspond to the parts of the rice-grain where the absorption of the light from lower and hotter levels is comparatively slight. In this way the appearance of the small, bright granules — unstable and in constant motion — could be accounted for. But thermal agitation of this sort would also give rise to ascending currents, towards whose summits condensation and the more intense emission of light would occur.

The granulated appearance of the photosphere — which is to be regarded as an intrinsic photospheric feature — is not the only evidence of the instability of the Sun's surface. Scattered over it are remarkable features of which the most obviously apparent are the spots and the faculae.

Spots and faculae. — The spots are dark areas that stand out boldly against the bright background of the photosphere. They are roughly circular in shape with, usually, jagged borders, and their dimensions vary over a very wide range, at times being large enough for the spot to become visible to the naked eye as a black dot on the solar disc. Though spots have undoubtedly been noticed in antiquity, and were well known to the Chinese astronomers of the Middle Ages, they did not seem to retain their interest for scientists. And it was not until Galileo had turned his first telescope towards the Sun that the spots were first described as features of the Sun itself — as local darkenings of its surface. This revelation demolished all the currently accepted ideas regarding the quality of pure and incorruptible fire that was supposed to be the Sun's.

Unlike the spots, which are darker than the photosphere, the faculae — vast, complex structures — are brighter. Thus they are best seen in the darker regions near the limb.

Spots and faculae are related to one another, for spots never appear without the prior formation of faculae, although faculae are frequently observed which have no associated spot. Let us look at these characteristics in a little more detail.

Generally speaking, a sunspot consists of a very dark umbra, surrounded by a fairly extensive lighter zone, the penumbra. The structure of the latter is noteworthy, for it appears to consist of numerous filaments converging on the centre of the spot, the whole thing giving the impression of some fantastic whirlpool. The accompanying illustrations give a better idea of the general appearance of spots than any amount of description; it will be seen that the structure of spot groups is extremely complex.

It is only in a manner of speaking that the umbra can be called black, or even very dark — a phrase which adequately describes its telescopic appearance — for its apparent darkness is solely an effect of contrast with the surrounding photosphere, whose brightness is four or five times more intense (Fig. 499). In reality, the umbra of a spot is more brilliant than an electric arc, and its real brightness can be readily demonstrated by comparing it with the absolute blackness of the Moon's limb when, as occasionally happens during an eclipse, it comes between the observer and a spot. A reddish tint is often to be noticed in the umbrae of certain sunspots, giving

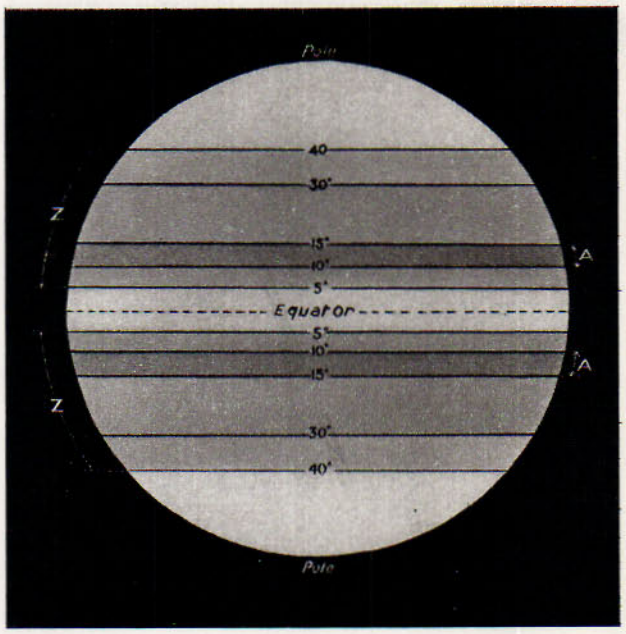


Fig. 505 — THE SPOT ZONES ON THE SUN

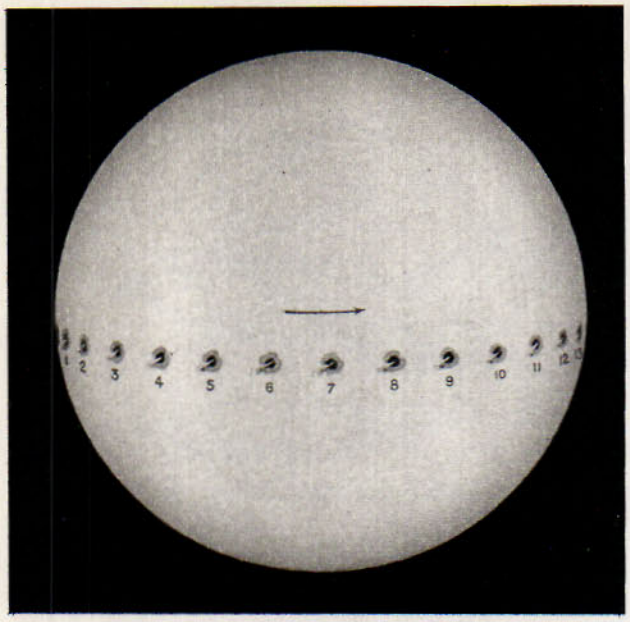


Fig. 506 — THE SUN'S ROTATION REVEALED BY A SPOT'S MOVEMENT ACROSS THE DISC

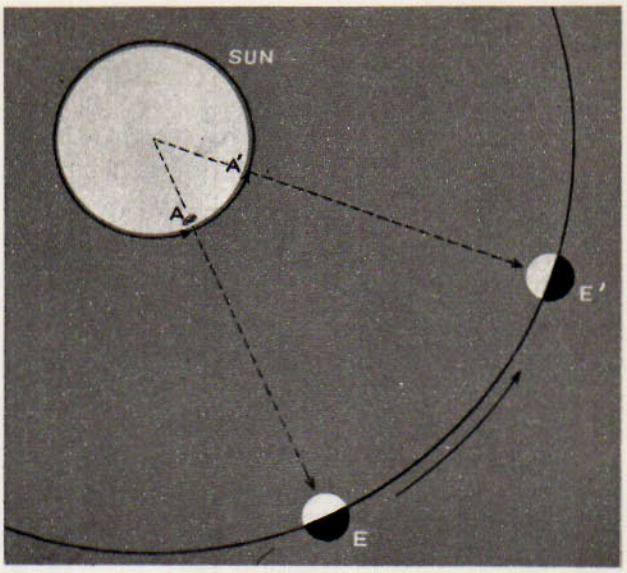


Fig. 507 — SIDEREAL AND SYNODIC ROTATION OF THE SUN

them a veiled appearance suggesting the presence of some relatively transparent absorbing material.

In short, the spots must be regarded as areas where the photosphere suffers local modifications of its general condition: in particular, a reduction of temperature. According to the best estimates, the temperatures of the spots are of the order of 4,000° C, i.e. 2,000° lower than that of the rest of the photosphere.

Then again, the spots are in no way stable or permanent features of the photosphere. Their birth, development, changes of shape, and ultimate disappearance, can be followed from day to day, though their durations vary widely — from less than a day for the smallest, to several months, sometimes, for the largest.

It seems that the birth of a spot is heralded by a disturbance of the photosphere, accompanied by faculae and eruptions of a type which we shall consider later, when dealing with the chromosphere; the mesh of the rice-grain pattern widens and tiny black spots (pores) appear in the interstices; these then develop into a greyish area, as though one were looking at a dark object through a veil; finally this 'veil' fades, leaving a spot or a group of associated spots whose area increases for several days, during which time they also undergo more or less marked changes of shape. As the life of a spot draws to its close, the photosphere appears to invade the penumbra, giving the appearance that it is contracting, while at the same time bright, bridge-like filaments form over the whole area; some time later nothing remains but a few pores and associated faculae. The photospheric disturbance then sometimes starts all over again, it being no rare thing

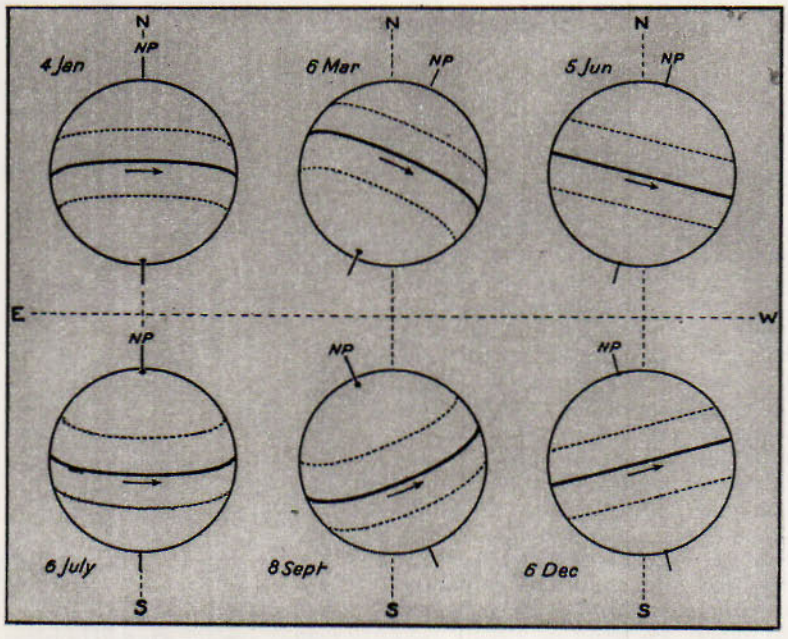


Fig. 508 — CHANGING VIEW OF THE SUN DURING THE COURSE OF ONE TERRESTRIAL YEAR

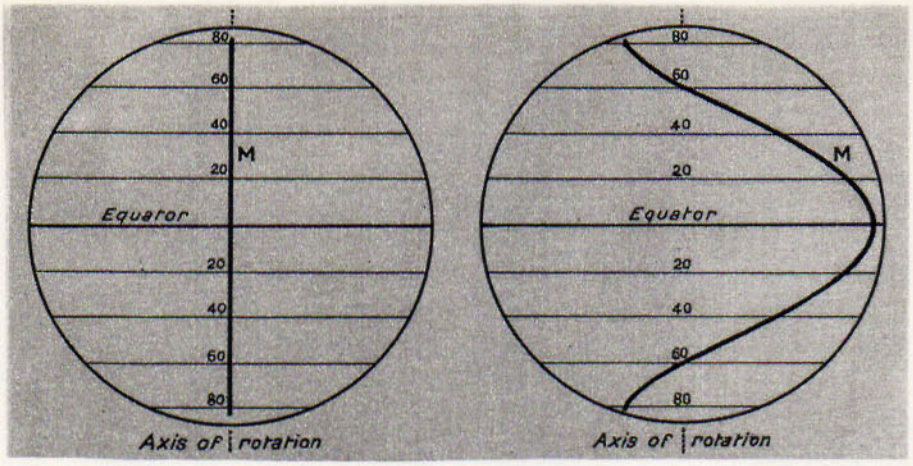


Fig. 509 — DISTORTION OF A MERIDIAN AFTER ONE COMPLETE ROTATION IN LATITUDE 60°, CAUSED BY THE SUN'S DIFFERENTIAL ROTATION

to see a new spot forming in the same position as an earlier one.

In general, these events do not occur at any definite positions on the Sun's surface. Indeed, the position of a spot during the course of its life is not fixed: concurrently with its changes of shape and size, its position also changes; the proper motions of spots, which are small or non-existent in latitude, are often very marked in longitude. In the case of spot groups the motions are rather complex. The layout of such groups is quite characteristic: they are elongated in a direction approximately parallel to the equator, and most commonly consist of two principal spots, which undergo dissimilar displacements along the line joining them. At first the western spot moves fairly rapidly forward along this line, covering perhaps 1° per day; subsequently it slows down, stops, and begins to retrogress slowly. The motions of the eastern spot are the reverse of this, though they are very much slower. It would thus appear that at the inception of a spot group there are forces in play which tend to separate the components violently, and that the motions thus produced are later damped down by the viscosity of the medium containing the spots. This statement does not seem to conflict with what is known of the state of the solar material, to which topic we shall return when discussing the peculiarities of its rotation.

Sunspots not only represent so profound a modification of the appearance of the photosphere that it can only be described as cataclysmic, but they also involve vast areas. Although some spots do not grow beyond modest proportions — several thousand miles in diameter — others attain truly impressive sizes. The diameter of any spot which is large enough to be seen with the naked eye, for instance, must be at least three times that of the Earth. If we imagine the umbra of such a spot as being a hole in the photosphere, allowing one to see into the Sun's dark interior (as was once thought to be the case), then our Earth could be swallowed up in it without touching its sides. Other spots far surpass even this, and when they are associated in groups the dimensions of the whole group may reach several hundreds of thousands of miles. Despite this, even when spots are most numerous, the total area of all such abnormalities of the photosphere does not exceed 1/500 of the visible hemisphere.

Even though the spots may not be holes in the photosphere, they are probably depressions. This is suggested by their changing appearance, the result of perspective, when viewed in successive positions around the Sun's globe (Fig. 503). From this it would appear that the umbra is the bottom of the depression, and the penumbra the slanting wall connecting it with the surface of the photosphere. But estimates of the depths of given spots have been extremely discordant — whereas it was once usual to speak of depths of 500 miles or more, it now seems to be well established that they are nearer 50 miles only.

The faculae, on the contrary, stand up in relief from the surface of the photosphere. Thus when they are seen at the limb they appear as fairly well marked projections. They are immense incandescent clouds, similar to those of the chromosphere.

We have learnt that spots are always accompanied by faculae, whereas the reverse is not true. This difference between the two groups of phenomena is further emphasised by the fact that faculae are to be found all over the Sun's surface, while spots are restricted to two zones between latitudes 5° and 30° in each hemisphere, though a spot may exceptionally appear as far from the

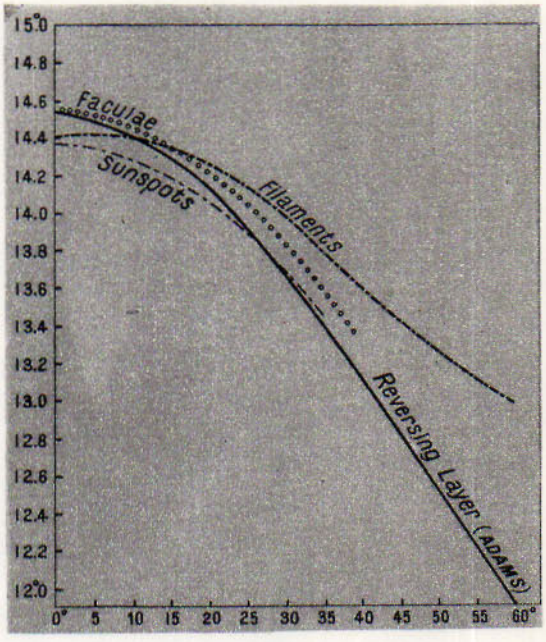


Fig. 510 — DIURNAL ANGULAR ROTATION AS A FUNCTION OF LATITUDE AT DIFFERENT LEVELS IN THE SUN'S ATMOSPHERE

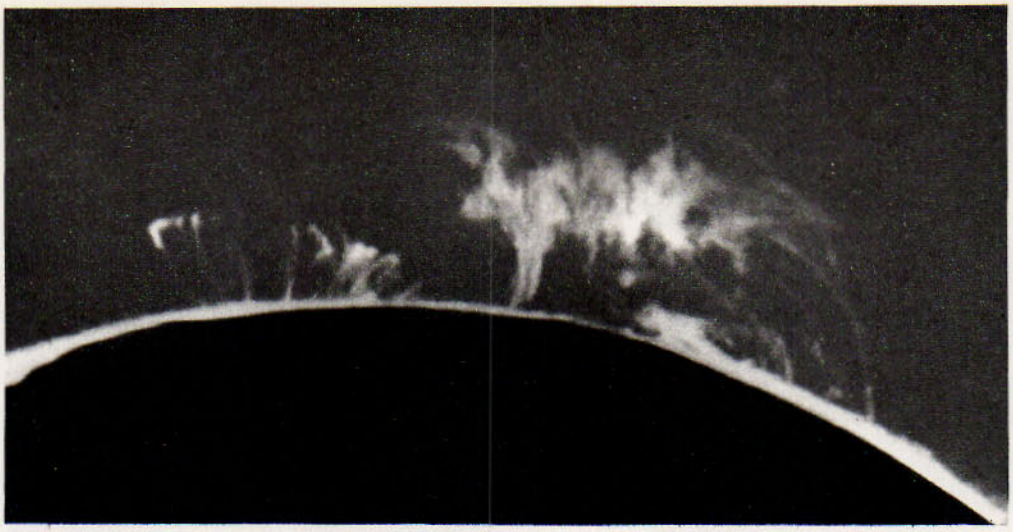


Fig. 511 — A FINE PROMINENCE ON THE SUN'S LIMB, 11 JUNE 1937. K_3 spectroheliogram. *(Meudon Observatory)*

equator as 40°, or even 45°. The rate of spot-production in these two zones is not fortuitous, but follows a well determined rhythm from maximum to minimum, in step with the solar cycle (described later). At the beginning of a new cycle, immediately following the minimum when there may be a fairly prolonged absence of spots, they begin to break out again on the high-latitude edges of the spot zones. As the cycle progresses, they gradually move in towards the equator. The spots situated nearest to the equator (the last of the old cycle) have hardly died out before the first of the new cycle begin to appear again in high latitudes; and so the process goes on. This law, describing the manner in which the distribution of the sunspots varies, is known as Spörer's law.

The Sun's rotation. — It is easy to establish, simply by observing the sunspots, that the Sun rotates from west to east; thus its surface features appear at the west limb and disappear at the east.

The Sun's axis of rotation is slightly inclined (7° 15′) from the perpendicular to the plane of the ecliptic, so that during the course of our annual circuit of the Earth's orbit we view the Sun, and the direction of its rotation, under a variety of slightly different perspectives. The line of intersection of the planes of the Sun's equator and the ecliptic (the line of nodes) is so oriented* that the Earth crosses it at the beginning of June and December. As Fig. 508 shows, the Sun is seen exactly in profile at these times, the spots describing straight lines parallel with the equator as the Sun's rotation carries them across the disc. At intermediate times the equator and the parallels of latitude appear more or less curved to the north or the south, as the result of the perspective of a spherical surface, while first one pole and then the other comes into view.

The observed motions of the spots indicate a mean rotation period of 27.25 days. But the period thus determined is the synodic rotation period, i.e. the period required for a given point to return to the same position

* The longitude of the ascending node, measured along the ecliptic, is approximately 75°.

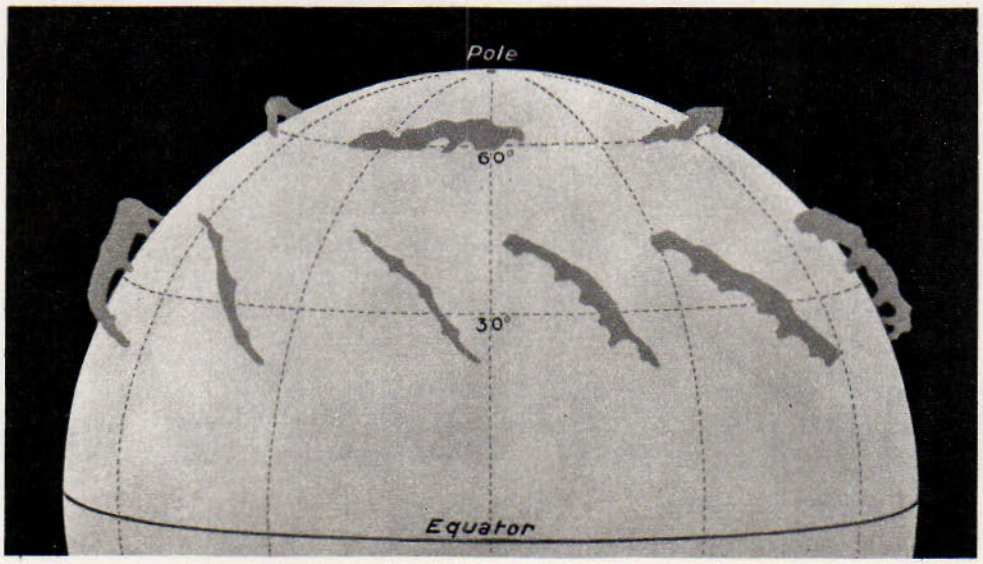


Fig. 512 — PROMINENCES IN HIGH AND LOW LATITUDES, AFTER L. AND M. D'AZAMBUJA

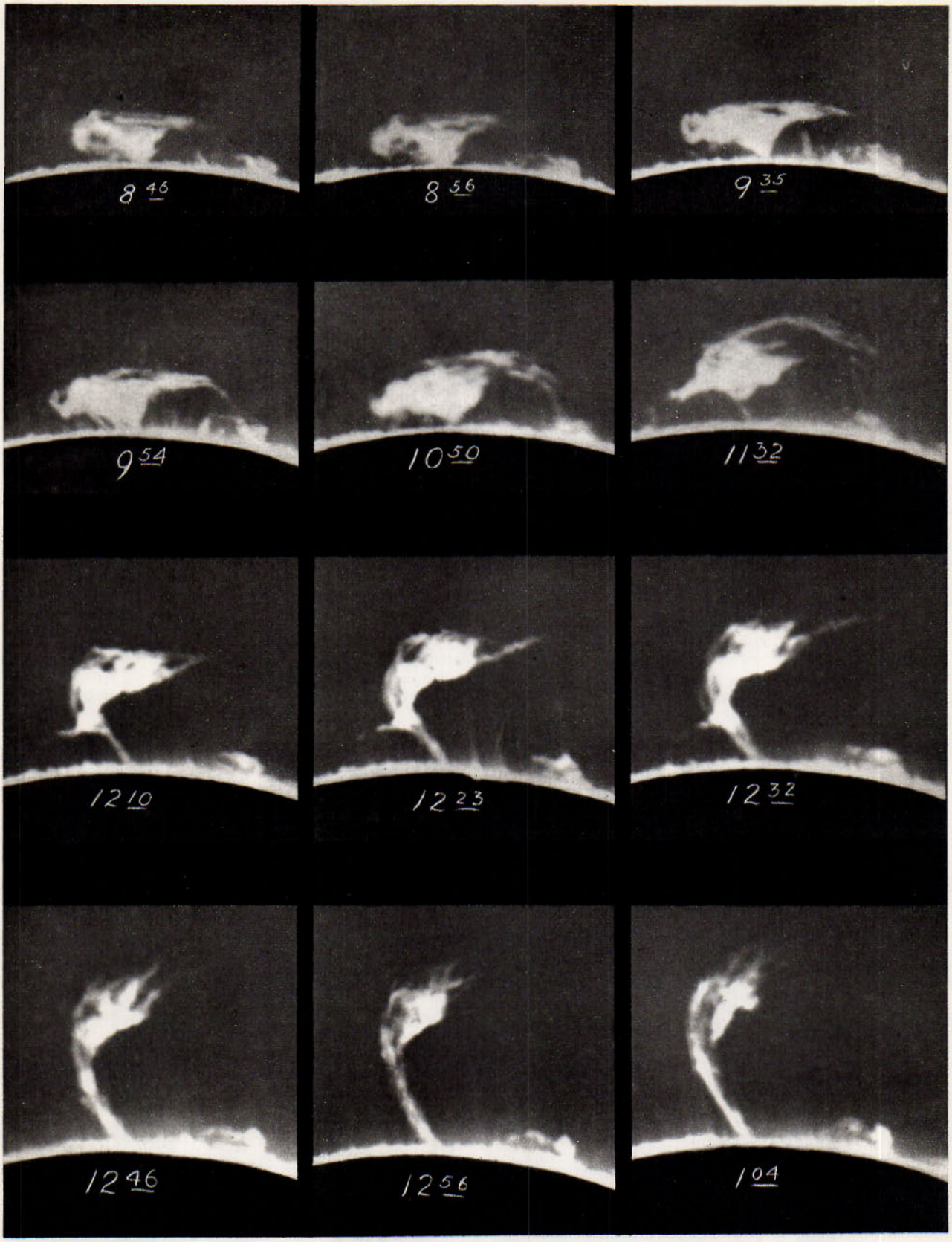


Fig. 513 — EVOLUTION OF AN ERUPTIVE PROMINENCE, 18 JUNE 1929. *(Photos, Mount Wilson Observatory)*

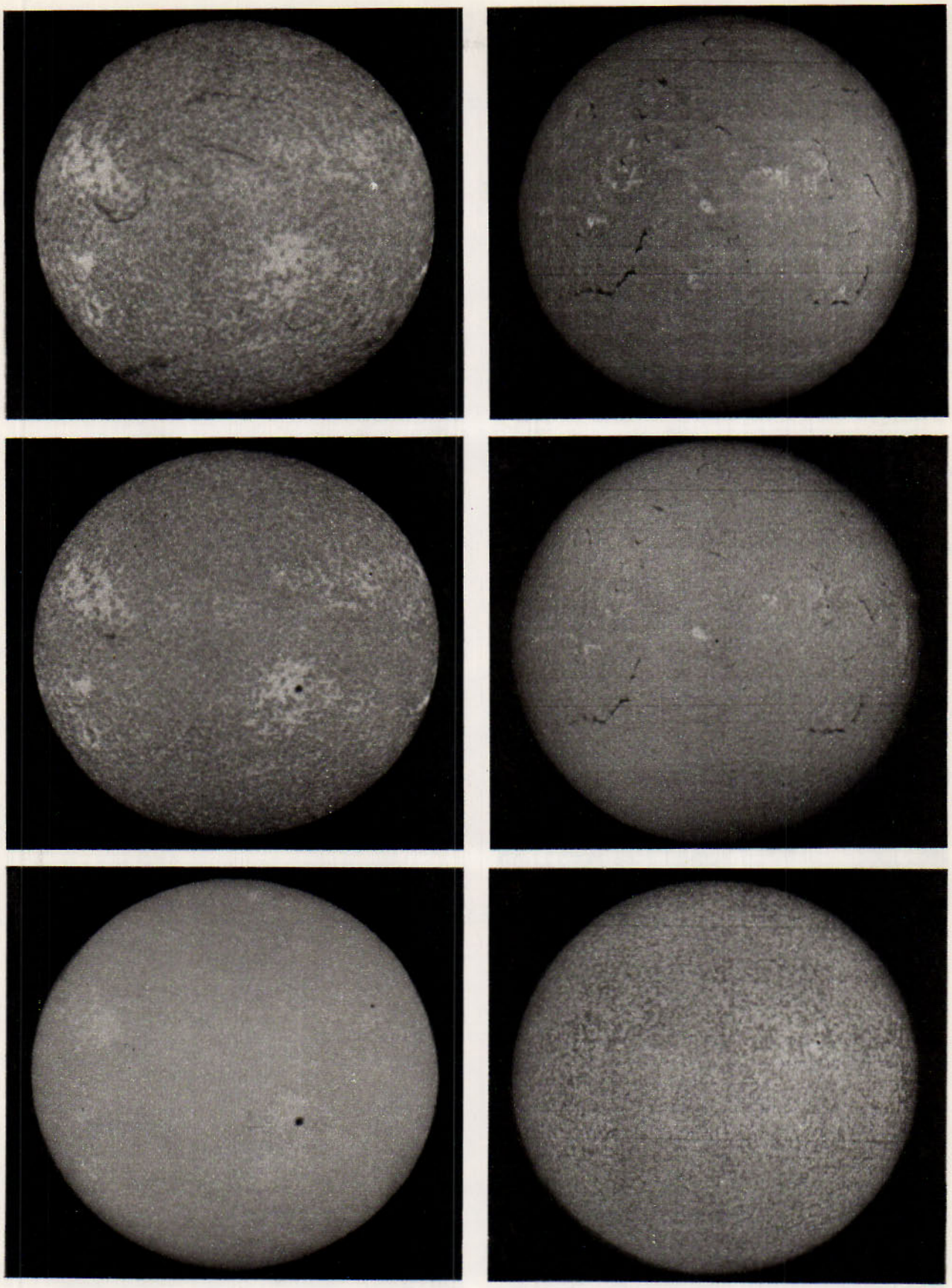


Fig. 514 — SPECTROGRAM TAKEN WITH THE K LINE OF IONISED CALCIUM (LEFT) AND THE Hα LINE OF HYDROGEN (RIGHT). These show, from top to bottom of the page, the upper, middle and lower levels of the chromosphere.
(Photos, Meudon Observatory)

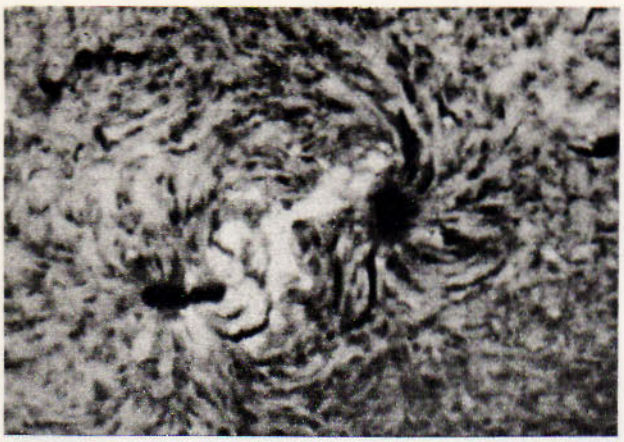


Fig. 515 — SOLAR VORTICES CENTRED OVER A SUNSPOT GROUP. 20 September 1926; Hα spectroheliogram. *(Photo, Meudon Observatory)*

on the disc as seen from the Earth. The distinction between solar and sidereal days (see p. 35) here crops up again: while the Sun is rotating, the Earth is also moving round it in the same direction (Fig. 507), thus changing the observer's viewpoint; when this is taken into account, we find that one mean sidereal rotation of the Sun is accomplished in 25.35 days.

Why 'mean' rotation? The answer is that the Sun does not — like the Earth, for instance — rotate as a solid body. The rotation period is a function of latitude, becoming progressively longer from the equator to the poles. We have already encountered a similar state of affairs when dealing with Jupiter.

The direct measurement of the Sun's rotation period by means of spot observations is not possible in latitudes higher than about 40°, and the rotation of the zones nearer the poles has been determined by means of the spectroscope (see Chapter 20), applied to the different regions of the Sun's gaseous envelope.

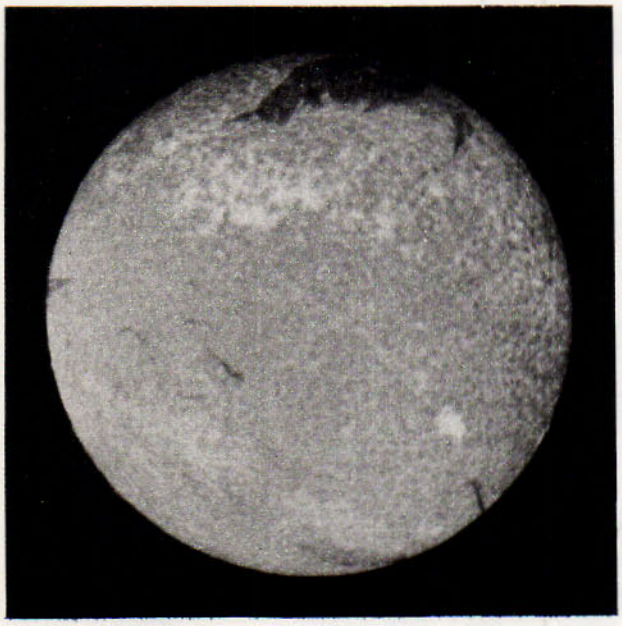


Fig. 516 — PHOTOGRAPH OF THE UPPER CHROMOSPHERE IN IONISED CALCIUM LIGHT, showing a very large prominence in the south polar region (top) as a filament projected on the disc. 2 June 1946. *(Photos, Meudon Observatory)*

By these methods, the following figures have been derived for the rotation period in different latitudes:

Latitude	Period (days)
0°	24.9
5	24.9
10	25.0
15	25.1
20	25.3
25	25.6
30	25.9
40	27.5
60	31.0
75	33.1
85	34.0

It follows from this that — since the solar rotation involves the slipping past one another of adjacent zones of latitude — the general rotation causes a torsion around the line of the poles, as is illustrated in Fig. 509. In this diagram an arbitrarily chosen meridian is taken to serve as a comparison line. After one complete rotation of a point in latitude 60° the more rapid rotation of the equatorial region, and the slower rotation of the polar, will have distorted this meridian in the manner shown; this deformation will be exaggerated with every succesive rotation.

This account does no more than sketch in the general outlines of the Sun's rotation. The observed phenomena are more complex, since the proper motions of the spots and other features seated at different levels are superimposed on the shearing effect of differential rotation. There is thus a combination of motions, which in certain cases complicates the determinations. This shearing effect between different zones of latitude, as well as probable variations of velocity with altitude above the Sun's surface, seems to provide a satisfactory explanation of the appearance and behaviour of phenomena whose spectacular development in the solar atmosphere is related to disturbances in the photosphere.

This leads us to the consideration of the various zones of the Sun's atmosphere.

Reversing layer, chromosphere and prominences. — According to the definition already given, the first atmospheric stratum to be distinguished above the level of the photosphere is the reversing layer, followed by the chromosphere. It must be emphasised, however, that the one merges insensibly into the other, and that one should think rather of successive zones with differing characteristics and phenomena, than of discrete layer-like geological strata.

The true significance of the reversing layer will not be appreciated until we have described the principles of spectroscopic analysis (Chapter 20), where the modifications that this layer causes in the radiation passing through it from the photosphere will be explained. For the present let it just be said that its spectroscopic properties permit us to identify on the Sun many of the same elements that make up the body and atmosphere of the Earth. Hydrogen and helium predominate, followed by oxygen, silicon and various other familiar elements.

The reversing layer, several hundreds of miles in thickness, is somewhat cooler than the photosphere. It may be regarded as part of the true atmosphere of the Sun, though an atmosphere of very great tenuity, for its pressure cannot be greater than 1/1000 that of the Earth's.

Above the reversing layer rises the chromosphere, so called on account of its bright rose tint. Without special equipment it is visible only during total eclipses, when it

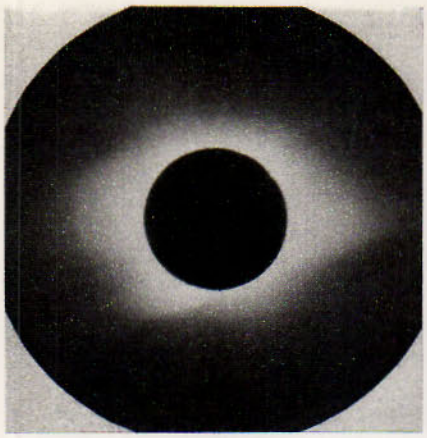


Fig. 517 — THE CORONA AT THE TOTAL ECLIPSE OF 1932

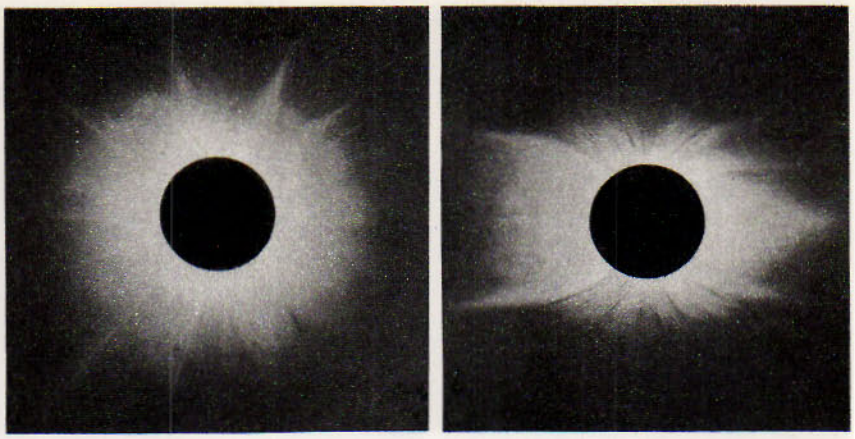


Fig. 518 — THE CORONA (SEEN DURING ECLIPSE) AT SPOT MAXIMUM (LEFT) AND SPOT MINIMUM (RIGHT)

appears as a narrow, bright ring surrounding the dark body of the Moon completely masking the photosphere. About 6,000 miles deep, it is rich in calcium, helium and hydrogen, to which it owes its colour. It is not of uniform depth, and its upper border bristles with small projections, giving it an appearance that has been likened to an immense field of grass.

Some of the most remarkable solar features are located in the chromosphere, among the more spectacular being the fantastic jets of gas which appear to spring from it: the prominences. Some of these, known as quiescent prominences, occasionally measure several hundreds of thousands of miles across at their base, though their height does not normally exceed some 15,000 to 30,000 miles. Sometimes they appear in the form of enormous clouds suspended in the Sun's atmosphere. They change only slowly, and sometimes survive for several weeks.

Very different are the eruptive prominences. Narrow at their base, resembling jagged flames, they are ejected violently from the chromosphere, often with velocities of 50,150 or even 250 m.p.s. They commonly attain heights of from 50,000 to 250,000 miles, though they have in exceptional cases been observed at heights of more than 500,000 miles — more than one solar radius. Their composition is similar to that of the chromosphere, often with the addition of metallic vapours. Some prominences, finally, appear white. They are extremely ephemeral, and subject to rapid transformations; often they disintegrate into separate spark-like components which vanish one by one while still rising from the chromosphere.

Like the chromosphere, the prominences can only be observed telescopically during eclipse; thus their study was originally limited to those few fleeting moments that are the astronomer's meagre ration of totality. But thanks to spectroscopic and photographic developments, and the application of associated physical techniques, these beautiful objects can now be studied at any time. Since their regular observation became possible, a great deal has been learnt regarding their true character and behaviour, as well as their relation to many other manifestations of the Sun's ceaseless activity. These new techniques will be described later, in the chapter devoted to methods of observation. They were initiated by Hale and Deslandres, who were the first observers to secure images of the whole chromosphere; more recently, B. Lyot has shown that these chromospheric phenomena may be photographed, and even recorded on movie film, with remarkable clarity and an astonishing richness of detail.

The data accumulated by these methods have revealed the numbers and distribution of these various features of the chromosphere. Periodic variations have been detected, and these can be precisely correlated with the variations in spot numbers, as a function of the solar cycle.

Prominences, like the faculae, are not restricted to the zones in which sunspots appear. They are noticeably more frequent between latitudes 20° and 40°, north and south, and these zones also move in towards the equator during the course of the solar cycle, dying out at sunspot minimum and reappearing in higher latitudes at the beginning of each new cycle. But there is a second maximum in their distribution, which, though it also starts at about 40° north and south at the beginning of each cycle, then moves progressively nearer to the polar regions, where it dies out during the paroxysms of activity that mark the succeeding cycle. It appears that these two distribution maxima must relate to prominences of different types, one of which is associated with the spots, while the others may develop in any part of the Sun's surface.

According to Pettit, in fact, it is necessary to distinguish not only between eruptive and quiescent prominences, but also between:

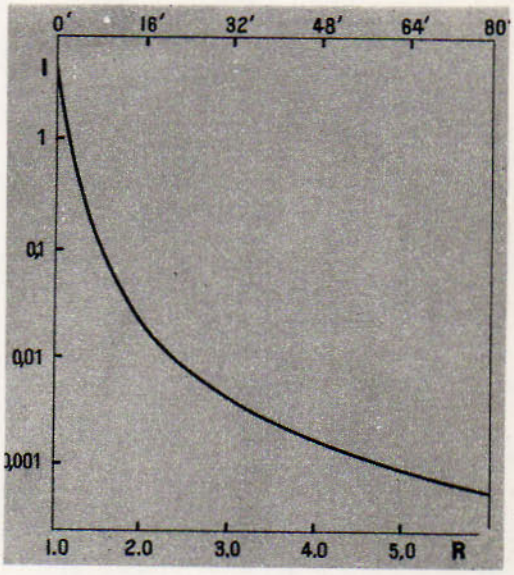


Fig. 519 — DECREASING BRIGHTNESS OF THE CORONA AS A FUNCTION OF DISTANCE FROM THE SUN'S LIMB (in units of 1 solar radius), measured during eclipses.

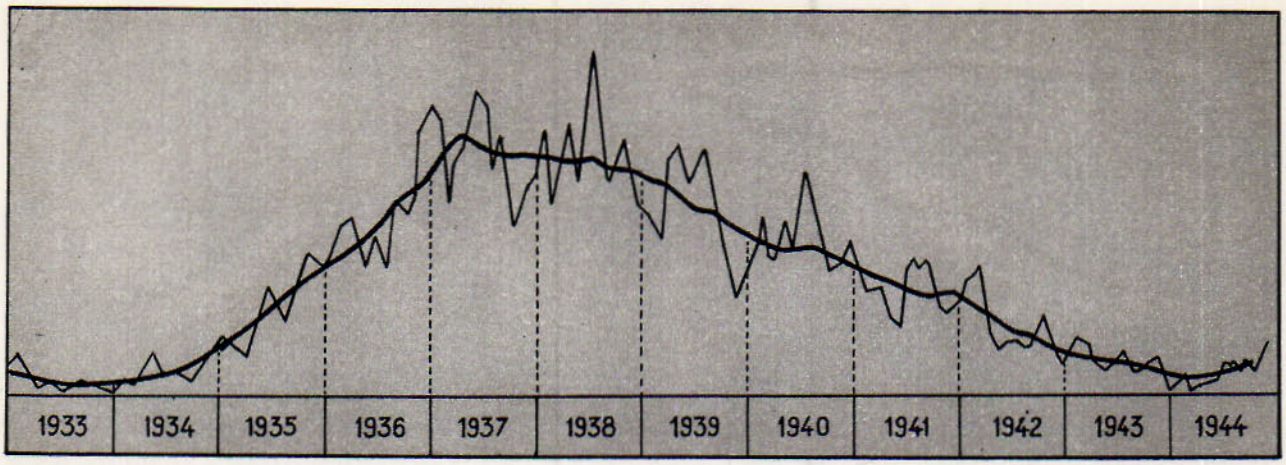


Fig. 520 — VARIATION OF SPOT NUMBERS DURING THE 1933—44 CYCLE

(a) Prominences whose form resembles splashes, which are always found in association with spots, around which they are distributed in a manner reminiscent of the lines of force in the vicinity of a magnet. It can be observed that the material of these prominences is moving in towards the spot centre.
(b) Spiral prominences, in the form of thin, bright, helical threads.

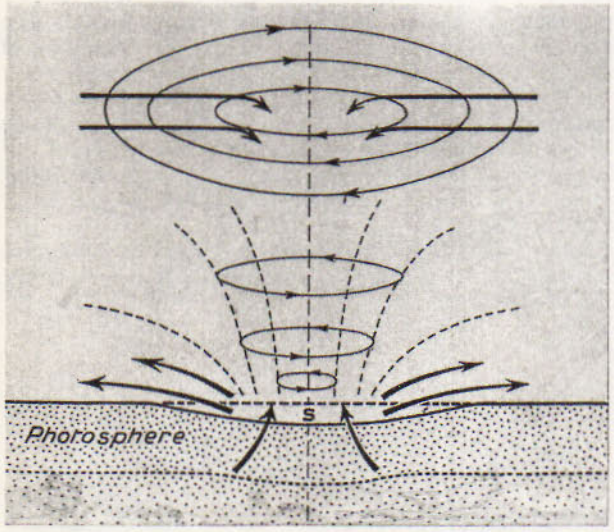


Fig. 521. Extremely diagrammatic representation of the structure of the chromosphere above a sunspot S, showing the motion of the gases (heavy arrows), the solar vortices (above), and the circular paths of the electrons (below), following the lines of force of the spot's magnetic field (dotted lines).

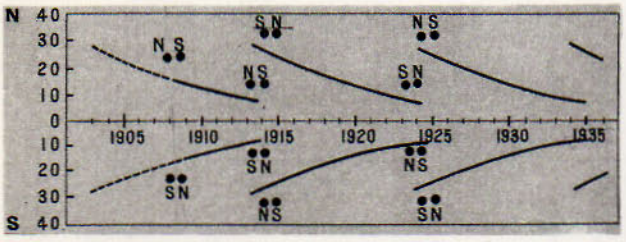


Fig. 522 — MEAN LATITUDES AND MAGNETIC POLARITIES OF SPOTS DURING THREE SUNSPOT CYCLES

(c) Eruptive prominences, subdivided into two groups; (*i*) jets of metallic vapours in the vicinity of spots; (*ii*) eruptions which develop and die out rapidly and are unconnected with the spots. Some of these have lateral motions which carry them back into the chromosphere again, tracing out a complete arc. Such motions have been strikingly demonstrated in some of Lyot's very beautiful accelerated movie films.

But these impressive solar show-pieces must not detain us too long: of equal interest is the detailed understanding of the chromospheric layers from which they spring. We have already mentioned the chromosphere's lack of uniformity. The distribution of its various constituent elements, and the manner in which they present themselves, vary considerably with altitude. In a general way it is convenient to consider the chromosphere as consisting of a lower, a middle, and an upper zone, each of which is illustrated in the typical photographs reproduced as Fig. 514, page 271.

It may be worthwhile considering these photographs for a moment. They are obtained (by the methods mentioned above) in such a way as to reveal only one particular chromospheric level, and show the distribution and structure of the calcium or hydrogen clouds (the principle elements in the chromosphere) at this level. If individual photographs of this sort are superimposed, the composite picture thus built up tells us something of the relations between the phenomena at the Sun's surface and those of the atmospheric levels that overlie it.

The elements composing the chromospheric layers are distributed in the form of bright clouds, to which the name flocculi has been given. Their non-uniform distribution produces a reticulated pattern which is at times reminiscent of the photospheric granulation, though on a much larger scale. But, most characteristically, the flocculi take the form of patches or clouds, extending over vast areas; it has been established that they develop above faculae, though they are larger than those. The middle levels of the chromosphere seem to be the most active, and it is here that the most notable features develop. The lower level, on the other hand, approximates closely to the photosphere in appearance and shares many of its characteristics.

Great clouds are to be observed in photographs of both levels — associated with spots and disappearing when they disappear. This association is also obvious from even a cursory examination, which shows that in general the flocculi are confined to two fairly definite zones, on either side of the equator, and that these coincide

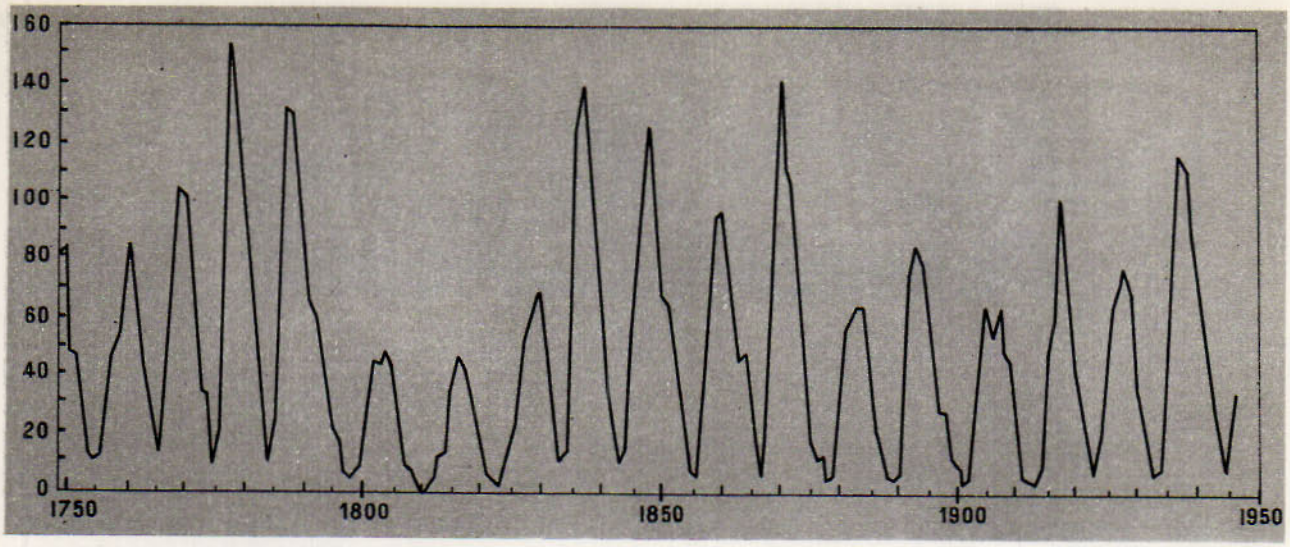


Fig. 523 — SOLAR ACTIVITY DURING 200 YEARS, ACCORDING TO STATISTICS COMPILED BY ZURICH OBSERVATORY

with the spot zones. Thus the incidence of flocculi, like that of spots, fluctuates with the 11-year cycle. On the other hand the fairly regular network of small flocculi represents a permanent feature of the chromosphere, which is to be observed covering the whole disc at all times.

These remarks apply equally to the upper levels of the chromosphere, though in certain details they are to be distinguished from those below. Hydrogen is here predominant, and both it and calcium are distributed in immense clouds which almost completely obscure the spots; such faculae as reach into this level are also more extensive. The clouds of hydrogen, moreover, assume a very characteristic whirlpool appearance, for which reason they are commonly known as solar vortices.

Photographs also reveal the existence of curious and sometimes immense dark filaments. Originally taken for a special type of feature, they are now known to be nothing more than prominences silhouetted against the brighter background of the disc. Lastly, there are the flares: intensely brilliant and usually very short-lived clouds which frequently develop in the chromosphere. We shall return to these later, when discussing the repercussions that solar events have on the Earth.

The corona. — Above the chromosphere the corona extends to immense distances from the Sun. Its general appearance at total eclipses is that of a fantastic halo of pearly-white light, fading insensibly into the darkness of the sky; only during the moments of totality can it be seen in its full glory.

Although it certainly extends to distances at least equal to a solar diameter (with tufts or plumes, known as coronal jets, sometimes extending to ten times this distance), no definite outer 'edge' can be attributed to it. The corona, itself extremely tenuous, simply fades into the near-vacuum of interstellar space.

Our knowledge of this remarkable outermost envelope of the Sun's atmosphere is still in a comparatively rudimentary stage. In the first place, although its existence was known of, it was for a very long time not recognised as an appendage of the Sun. Although it is only visible to the naked eye during total eclipses, it contributes such an unprecedented splendour to these that it has certainly been known from the earliest times. Nevertheless, the first extant description of it was given by Kepler (who observed it during the eclipse of 1605); the next, that of Cassini (eclipse of 1706). At that time, however, it was attributed to our own atmosphere, and its recognition as part of the Sun had to await the observations of Baily and Arago at the eclipse of 1842.

The occasions on which the corona can be observed remain extremely rare, and the aggregate of the periods of totality in any year amount to only a few minutes. Whereas the lower layers of the Sun's atmosphere have, since the end of last century, been accessible to study at any time, this delicate outer envelope for a long time stubbornly resisted all the attempts of astronomers to observe it outside eclipse. It was less than 30 years ago that a brilliant extension and refinement of existing observational techniques enabled the French astronomer Lyot to record it at other times than total eclipse. Even so, he could only observe the brighter regions in the vicinity of the chromosphere. Despite this restriction, the investigations opened up by this development have thrown much new light on the many problems connected with the corona, about which practically nothing detailed was known.

Although the corona was only observable visually during the brief moments of totality, these moments slowly added up during the course of years, and it became evident that its size and shape are subject to large-scale changes; also, that these variations are closely connected with the rise and fall of the spot cycle.

The corona has a rayed structure, whose local characteristics and individual details are related to the form and extent of the corona as a whole. At times of maximum solar activity it assumes the form of a magnificent aureole, roughly circular and somewhat reminiscent of a dahlia, sometimes with numerous and irregularly distributed petals; other details that may be observed are dark arches spanning any prominences that happen to be visible at the time. At spot minima, on the other hand, the corona is concentrated rather in the plane of the equator, with its principal rays considerably elongated in this direction; towards the poles, shorter, curved rays — suggestive of the lines of force in the neighbourhood of the poles of a magnet — may be seen. Finally, at times intermediate between maximum and minimum the form of the corona is likewise intermediate between those just described: the longest rays develop more or less opposite the spot zones, giving the corona a roughly rectangular appearance.

These changes of shape originally led astronomers to

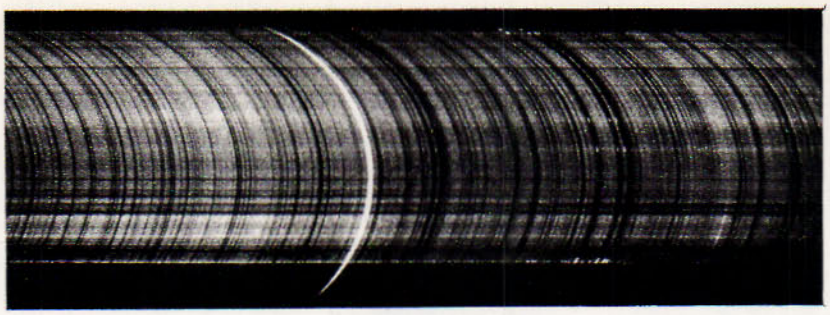


Fig. 524 — THE GREEN CORONAL LINE AT λ5303. The curvature of the lines results from the use of a curved slit, applied to the solar limb. The numerous dark lines are caused by diffusion of the Sun's light in our atmosphere, while the horizontal dark lines are due to particles of dust between the jaws of the slit of the spectrograph. The bright bands and luminous points are spectra of prominences cut by the slit. On the right, the weak coronal line at λ5116.
(Photo by Lyot, Pic du Midi Observatory, 18 July 1937)

believe that the matter composing the corona must be in constant motion. The too brief moments of totality hardly favoured the investigation of these motions, but some results were nevertheless obtained by successive observations at short intervals from different stations along the totality zone of a single eclipse. These at least proved that rapid changes do not occur in the corona. The classic investigations of Lyot — made possible by his development of a technique for the continuous recording of the inner corona — not only provided a rich yield of detailed information regarding its structure and nature, but also effectively verified the existence of internal motions whose slowness is in marked contrast to the violent turbulence of the chromosphere.

The solar cycle. — It will have been made clear by the foregoing descriptions of the Sun's varied features, and the changes they undergo, that the activity of this immense furnace fluctuates in a regular manner. At times held in the grip of vast seismic upheavals, at others relatively calm, the Sun might fancifully be likened to a patient suffering from a recurrent fever. It is the general character and progress of this fever that we must now consider.

The difference between the maxima and minima of the Sun's activity is very marked, as is evident from even a casual study of the sunspots. Whereas at maximum spots occur in large numbers and cover wide areas of the surface, at minimum the Sun may be completely free from them.

That the Sun's activity is subject to wide fluctuations was recognised as soon as its effective observation with the telescope began; but it was not until 1843 that Schwabe, a German amateur astronomer, discovered the true character of the sunspot cycle. This cycle, once discovered, became the object of intensive investigation.

If a curve is plotted of the daily number of spots visible (or of their total area, expressed as a fraction of the visible hemisphere) it will be found that this index of the Sun's activity contains recurrent and well-marked maxima, separated by equally definite minima. The general trend of the curve is such that the lapse towards minimum is slower than the rise towards maximum. The mean interval separating two maxima is 11.1 years, of which the fall to minimum occupies 6.5 years, and the rise to maximum 4.6. It is necessary to speak of the 'mean' period, since its duration is itself variable, sometimes stretching to as much as 17.1 years (the period 1788—1805) or shrinking to as little as 7.3 (the period 1830—7). We shall note other peculiarities of the solar cycle in due course.

Successive maxima are not necessarily of equal importance. Nor are they always sharply defined, for a secondary maximum may follow the first after a varying lapse of time (as, for instance, after the maximum of 1905). The minima are similarly of somewhat unpredictable durations, and they may also be enlivened by the unexpected appearance of spots, sometimes of considerable size.

Despite such minor irregularities, however, the periodic variation of the Sun's activity is a fact whose importance and significance cannot be over-emphasised; its repercussions, as we shall see later, are even felt on our own Earth. The cause of this well-established periodicity is, however, still unknown.

It has been suggested that other periods, of greater or smaller durations, may be superimposed on the fundamental 11-year period, and that these may account for the latter's irregularities Investigations into this possibility have not as yet produced any conclusive results. Statistical studies (notably by Waldmeier and Gleissberg) have in recent years nevertheless isolated certain general properties of the cycle which to a limited extent permit the phases of the solar activity to be forecast. Nevertheless, it has long been recognised that there is a periodicity in the Sun's activity whose duration is longer than the 11-year cycle.

The physical study of the spots has revealed that they are the centres of intense electromagnetic disturbance. As long ago as 1928, Hale, at Mount Wilson Observatory, demonstrated the existence of magnetic fields above the spots, identical in appearance with that in the vicinity of the poles of a magnet. The average intensity of these fields is about 2,000 gauss, occasionally exceeding 4,000 gauss; for comparison, that of the Earth's field is 0.5 gauss. Spots in the same hemisphere have the same polarity, but this polarity is reversed from cycle to cycle. In other words, the polarities of the spots differentiates sharply between consecutive cycles, each with a mean duration of 11.1 years. These should thus be regarded as the two halves of a cycle of about 22 years, and the latter period certainly represents the true cycle of solar events.

Nature of the Sun's atmospheric layers. — We have seen that all the observable regions of the Sun are involved in the 11-year cycle of activity: events in the photosphere, the chromosphere and the corona, however different they may be, nevertheless march strictly in step with one another. These phenomena have been described, and need not be referred to again except in connection with the physical conditions in the layers that gave them birth.

We have also summarised our present knowledge of

the photosphere — the solar level, with its curious granular structure and temperature of 6,000° C, from which comes practically all the light which we receive on Earth. Above it lies the reversing layer: its temperature is about 5,000° and its low density becomes even lower with increasing altitude. Above this again is the chromosphere, with a thickness of some 7,000 miles, throughout the whole extent of which most of the same elements are encountered.

The physical conditions in the chromosphere are markedly different from those of the lower levels. It is in fact a region of such extreme rarefaction that in order to give a preliminary idea of it, it may be likened to a 'vacuum' such as not even the most refined laboratory techniques can emulate. It is not a gas, in the ordinary meaning of the word, but rather a region of isolated atoms maintained in suspension by selective radiation pressure — a process that has already been described in connection with comets (see Chapter 8). The same fundamental principles are involved in an explanation of the chromospheric environment and of the formation of comets' tails.

In the region nearest to the photosphere, which contains heavy elements (metals), conditions approximate to that of an atmosphere in thermodynamic equilibrium. Its upper limit, however, is difficult to define; between about 600 and 1,000 miles above the reversing layer there is, indeed, a transitional zone in which the action of light pressure begins to make itself felt on certain elements, and whose thickness therefore varies according to the nature of these elements. Above this lies the chromosphere proper, where hydrogen, helium and ionised calcium in radiative equilibrium predominate; at this level, light pressure has assumed the paramount rôle. In order to form some idea of the conditions in the chromosphere, one may liken its suspended atoms, supported by light pressure, to the pingpong balls that dance on jets of water in a fairground shooting gallery.

This, however, can only be regarded as a very general picture. The impalpable substance of the chromosphere can remain in equilibrium only so long as the temperature of the solar surface remains constant — which in fact it is far from being. Only consider the spots and the faculae: the former are relatively dark and cool, while the latter are hotter than the general photospheric surface; radiation pressure is therefore less intense above a spot, and a local collapse of the chromosphere will occur — its atoms behaving as though they were being attracted towards the spot. Conversely, the intense radiation of the faculae repels these elements to greater heights. As a result, turbulence is set up between the lower and upper levels, which is further complicated by motions in other directions resulting from the fact that the radiation from below does not necessarily traverse the chromosphere directly, but from every possible angle.

The continuous mixing produced by these motions justifies us in picturing the composition of the chromosphere as reasonably homogeneous, its atoms floating in the flood of photospheric light. Theoretical considerations, however — reinforced by our knowledge of downward motions of the gases of the chromosphere — appear to indicate that the unaided action of radiation pressure is insufficient to maintain this equilibrium; in short, the chromosphere ought to collapse slowly towards the Sun's surface. That it does not do so must be due to the intervention of other factors, such as the violent eruptions that produce prominences: these constantly reinforce the upper chromosphere. This question is, however, unfortunately too complex to follow up in detail in a general account of this sort.

From the chromosphere we pass on to the corona,

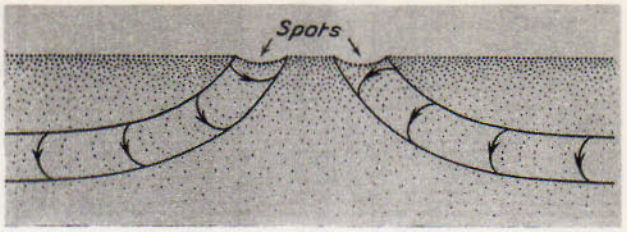


Fig. 525 — VORTICES GIVING BIRTH TO SUNSPOTS, ACCORDING TO BJERKNESS'S THEORY

which again represents a very different type of environment, displaying characteristics entirely peculiar to itself. If the chromosphere can still be thought of as a gaseous envelope — at any rate so far as its lower levels are concerned — the same cannot be said of the corona. For if it were, the high pressures resulting from its immense extent would confer on the chromosphere a density which in fact we know that it does not have. Another proof of the extreme tenuity of the corona is furnished by the fact that when comets pass through it at high velocities their motions suffer no modification whatever. Nothing of which we have first-hand knowledge on the Earth can give us any inkling of the degree of tenuity of the corona.

Our understanding of the precise nature of this beautiful outer envelope of the Sun's atmosphere is considerably less advanced than that of the levels nearer its surface: we must remember that its effective study is restricted to the rare and brief moments of totality. Certain interesting conclusions have nevertheless been established.

The greater part of the light of the corona is photospheric; the manner of its diffusion and degree of polarisation suggests that the corona is composed of free electrons. By reason of the excessive tenuity of the whole structure, the most plausible theory would in fact seem to be that which treats the corona as consisting essentially of electron gas. Its rayed structure (equatorial petals and

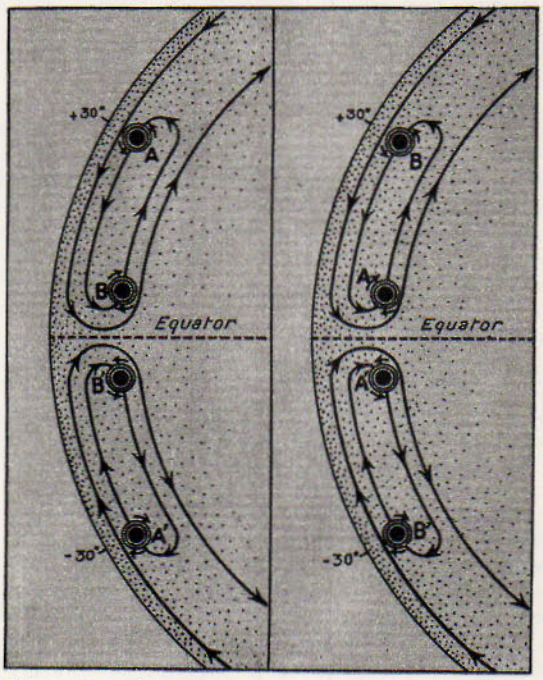


Fig. 526 — CIRCULATION OF VORTICES DURING THE COURSE OF 23 YEARS, ACCORDING TO BJERKNESS'S THEORY

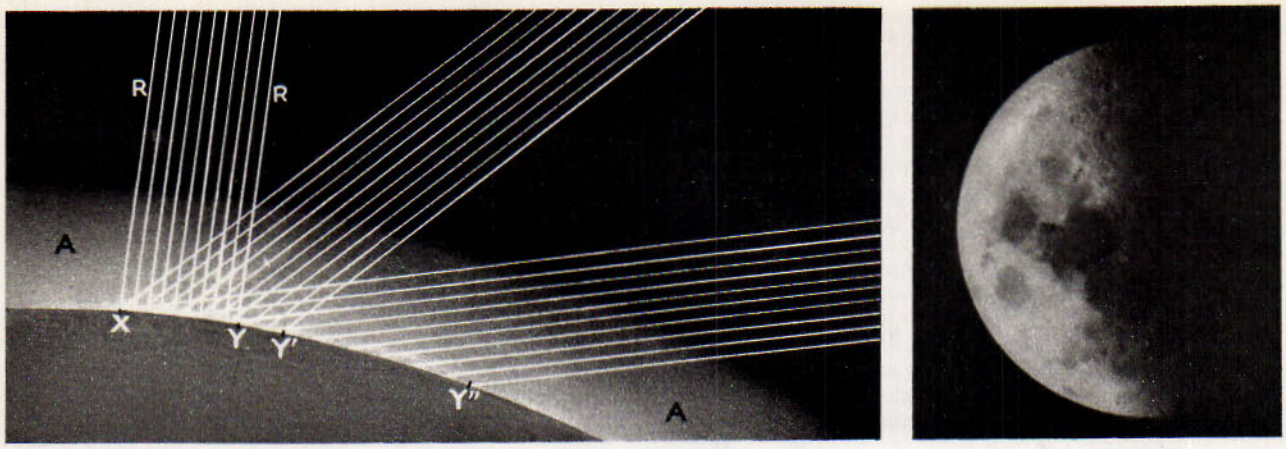


Fig. 527 — Left: a bundle of solar rays of given diameter is increasingly concentrated at the Earth's surface as its inclination to the horizon is increased. Right: the gradual darkening of the Moon's disc towards the terminator shows clearly the decrease in illumination as the angle of incidence of the rays increases.

polar filaments) would mark the trajectories of the electrons under the influence of the Sun's general magnetic field (less than one gauss) which was discovered by Hale and confirmed by more recent work; the variations of form, following the 11-year cycle, would thus be explained by the varying rate of emission of electrons from the spots and the active regions associated with them.

But in addition to diffused sunlight, the corona emits radiations of its own. Astrophysicists were for a long time puzzled by these, for when they were analysed spectroscopically they could not be identified with those emitted by any of the known chemical elements: they were accordingly attributed to a mysterious element, unknown on Earth, to which the name 'coronium' was given. However, there are no gaps in the periodic table of the elements into which such a substance could be fitted, and it was eventually realised that what we have here is a known (though unidentified) element or elements radiating under abnormal physical conditions.

The enigma could, indeed, only be solved by invoking special physical conditions, under which a familiar element is able to emit radiations which are prohibited under the experimental conditions which can be produced in our laboratories. In the language of modern physics, we say that the atoms are several times ionised — i.e. they have lost several of their electrons. Later on, we shall describe in greater detail the conditions required to produce such a state; it will suffice to say here that, given extremely low densities, the atoms can emit so-called 'forbidden' radiations. By this is meant the radiation of wavelengths that cannot be produced experimentally even with the help of all the resources of modern laboratories. The mechanism of their production being the discovery of the theoreticians, observational verification consisted in determining if the observed wavelengths in fact agreed with the predicted ones for any given element.

The solution of this problem is of fairly recent date. In 1941 the Swedish astronomer Edlen, working from experimental data, succeeded in showing that the mysterious coronal radiations are produced by the atoms of ordinary metals, notably iron and nickel: it is only their physical environment that is unusual. Under the maltreatment that they receive in such an environment, the iron atoms are stripped of 9 to 12 electrons, and the nickel atoms of 11 or 14. Calcium is a third element that has been detected under like conditions.

Thus another mystery of astronomical spectroscopy went the same way as that of 'nebulium' (identified by the American astrophysicist I. S. Bowen in 1927), after it had baffled the most eminent astronomers and physicists since the beginning of the century.

But it must not be imagined that the enigma of the solar corona has been completely cleared up. Indeed, the presence of such highly ionised atoms (which have lost about half their normal complement of electrons) is extremely difficult to explain in the light of the known facts, for it would seem to indicate that the temperature of the corona is incomparably higher than that of the Sun's surface — several thousand million degrees, as compared with 6,000°; furthermore, this extreme temperature seems to be required in order to explain other features of the coronal spectrum, such as the widths of its lines, the absence of the Fraunhofer lines from the continuous spectrum of the inner corona, and the absence of the Balmer series of hydrogen lines — all of which imply an extremely violent thermal agitation of the coronal material. No generally acceptable alternative explanation, however, has yet been proposed.

One fairly plausible hypothesis supposes that the material of the corona may issue from the Sun's interior, where the temperature is a million degrees or more; immense jets of gas at high temperatures (hence invisible at the moment of their expulsion) are envisaged as leaking out through cracks in the photosphere, probably in the vicinity of spots. These expelled gases, whose atoms will already be stripped of many of their electrons, will cool slightly, as a result of expansion and radiation, and their atoms will gradually recombine with the free electrons which form the greater part of the coronal material, till finally they emit the characteristic radiations whose origin we have just been discussing.

Many details of all these fascinating problems remain to be cleared up: the magnificent glowing halo that surrounds the Sun presents a vast field for investigation, and one which astronomers and physicists are still far from exhausting — as can be said, with equal truth, of the actual body of the Sun itself.

Nature and origin of the photospheric features. — In appearance, the spots are suggestive of immense eddies or whirlpools, and this conception is strengthened by what we have just learnt about the motions and turbulence in the upper levels of the atmosphere.

In 1865 Faye propounded a cyclonic theory of the sunspots which envisaged them as vast whirlpools rotating about a vertical axis. The fundamental principle, at least, of such a theory seems to be supported by the spiral arrangement of the filaments comprising the penumbrae,

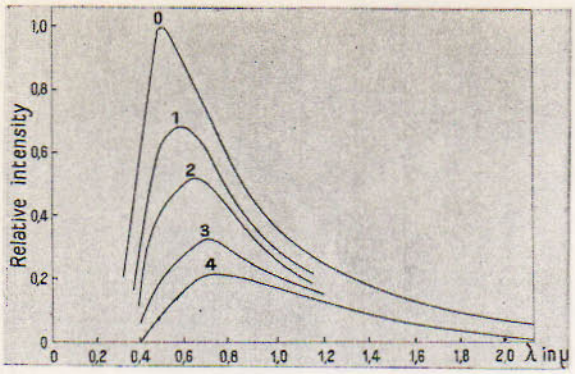


Fig. 528 — WAVELENGTH DISTRIBUTION OF THE SUN'S ENERGY AS A FUNCTION OF ITS ZENITH DISTANCE:
0 solar energy outside the Earth's atmosphere
1 Sun at the zenith
2 Sun 60° from the zenith
3 Sun 76° from the zenith
4 Sun 81° from the zenith

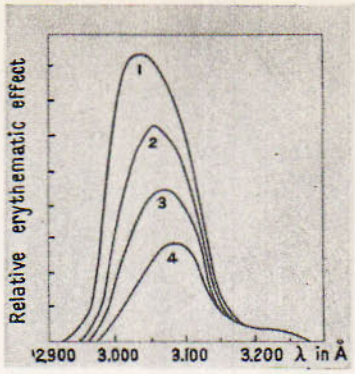


Fig. 529 — ERYTHEMATIC (SUNBURN) EFFECT OF THE SUN AS A FUNCTION OF ITS ZENITH DISTANCE:
1 Sun at the zenith
2 Sun 60° from the zenith
3 Sun 76° from the zenith
4 Sun 81° from the zenith

and by the more recently discovered hydrogen clouds in the upper chromosphere, whose vortical structure is of the same kind.

However, there were many observational facts which Faye's theory could not account for satisfactorily — the occurrence of irregular and broken-up spots, the character of spot groups, and many details of spot structure. Alternative theories were suggested, but these need only be glanced at in passing since they are now of no more than historical interest.

Thus the possible effects of the fall of meteor swarms, revolving approximately in the plane of the ecliptic with a period of 11 years, were discussed. This theory could account for the Sun's equatorial acceleration, but not for the distribution of the spots and their progressive approach towards the equator as the cycle advances.

The coincidental similarity of the duration of the solar cycle and the revolution perid of Jupiter (11.86 years) was likewise not overlooked. It was suggested that this giant planet may produce some sort of 'tides' which in turn cause the formation of spots. But the actual mechanism of such a process was left in doubt.

Another school of thought regarded the spots as evidence that the Sun is cooling. This, however, is contradicted by the results of recent work on the evolution of the Sun, which show that, on the contrary, the Sun's temperature in the future will be higher than it is now (see Chapter 18).

In an attempt to explain the fundamental causation of the phenomena of the photosphere, the Norwegian physicist Bjerkness developed a theory which was firmly based on the observational facts. According to this theory a sunspot is a whirlpool in the photosphere which produces a depression in its surface and sets up a suction effect which tends to draw matter up from lower levels. The upward motion of the incandescent gases results in a reduction of their pressure, with consequent expansion and a fall in temperature: hence the reduced brightness of such regions. The theoretical figures agree (at any rate as regards the order of magnitude) with the actual measurements of spot temperatures.

The expanding gases would be expected to spread

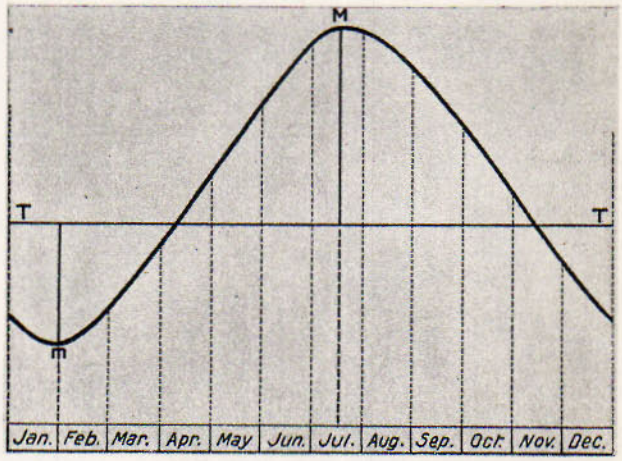


Fig. 530 — ANNUAL VARIATION OF THE TEMPERATURE OF THE EARTH IN TEMPERATE LATITUDES

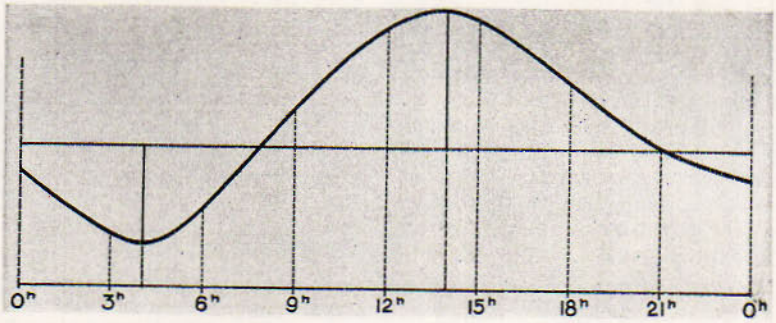


Fig. 531 — MEAN DIURNAL VARIATION OF THE TEMPERATURE OF THE EARTH IN TEMPERATE LATITUDES

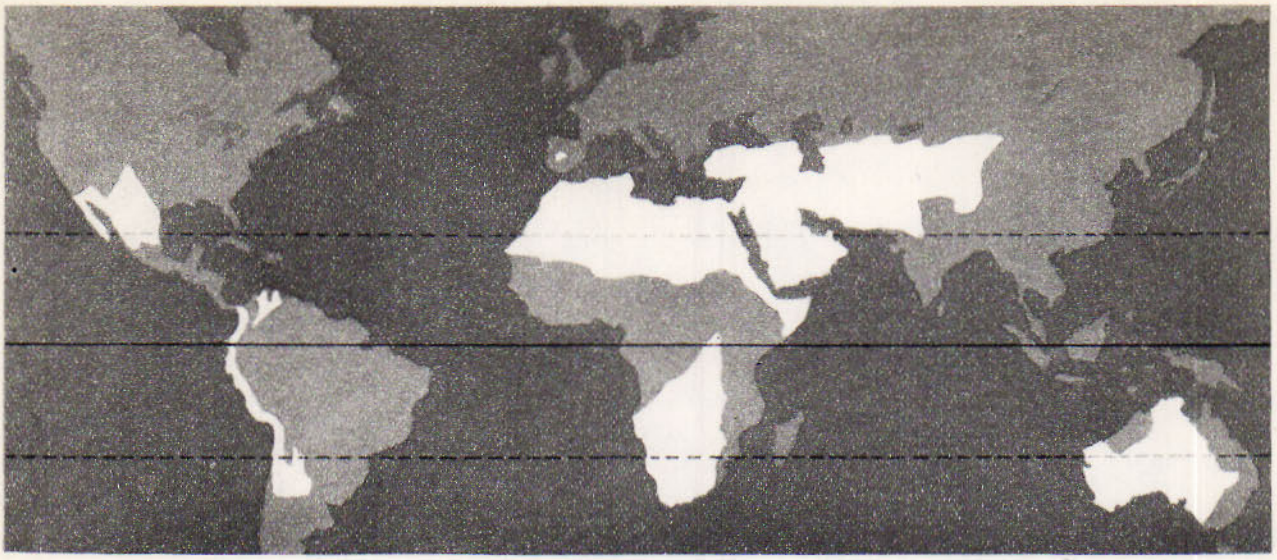


Fig. 532 — REGIONS OF THE EARTH IN WHICH THE UTILISATION OF SOLAR ENERGY IS A PRACTICAL PROPOSITION

outward across the surface of the photosphere, giving rise to effects which have in fact been confirmed by the spectroheliograph. Then, as has already been described, the reduction of the intensity of the radiation will disturb the equilibrium of the overlying layers, whose material will relapse towards the solar surface, their replacement setting up other currents as material flows in from the neighbouring regions.

The whole structure nevertheless lacks the coherence of a simple, isolated vertical whirlpool. The motion of the upper chromospheric hydrogen appears to be only indirectly associated with the photospheric vortex that forms the sunspot, for their directions of rotation are not necessarily the same. And the further fact that in the vast majority of cases the high-level vortices in each hemisphere rotate in the same sense as terrestrial cyclones suggest that they are influenced in some way by the Sun's rotation.

But however intriguing these suggestions regarding the nature of the spots may be, it must be remembered that they are only hypotheses. And the theories that seek, in addition, to account for the origin of the spots are in even worse case, owing to our ignorance of the real condition of the Sun's material, and particularly of the physical conditions reigning in its interior. It is nevertheless possible to appeal to certain observational facts in support of an attempt to correlate spot formation with the Sun's rotation. The theory of the German astronomer Emden was framed along these lines. He considered the Sun as divided into a number of different layers, each rotating with a different angular velocity; vortices set up by the friction between these successive strata would result in the birth of spots. His theory could not, however, give any account of the 11-year cycle.

Let us return to Bjerkness's theory. This envisaged the existence in each hemisphere of arch-shaped vortical zones, or tunnels, in the Sun's interior which rotate along almost identical parallels of latitude. The sections of these vortices, at the points where they emerge from the surface of the photosphere, are the sunspots: it will be remembered that the latter usually occur in pairs, the members of which have roughly the same latitude. In this way, also, the opposite magnetic polarities are explained, since the two sections of the tunnel, as seen from our exterior viewpoint, would rotate in opposite directions.

The periodicity and progressive latitude displacement of the spots is explained by their participation in a general circulation of the solar material from the poles towards the equator, and back again towards the poles. In each hemisphere this circulation would be as follows: surface material would descend from the pole towards the equator, losing heat by radiation, and would then return towards the pole at a deeper level, in the interior of the Sun, where it would regain its original temperature. And if one imagines that the two vortices A and B (Fig. 526), rotating in opposite directions, are involved in this circulation, whose period is 22 years, one can see that a spot cycle will begin when A arrives at the surface in latitude 30° to 40°, whence it will move towards the equator and die out some 11 years later; then B in turn comes to the surface and initiates a new cycle; and so the process is repeated. Since the direction of rotation at the two ends of the vortical tunnel is opposite, a change of spot polarity would be expected with each new cycle, as also the opposite polarities in the two hemispheres which are in fact observed.

This hypothesis is undoubtedly both ingenious and attractive, but we are still left in the dark as to the cause of the general circulation, its speed (which sets the pace

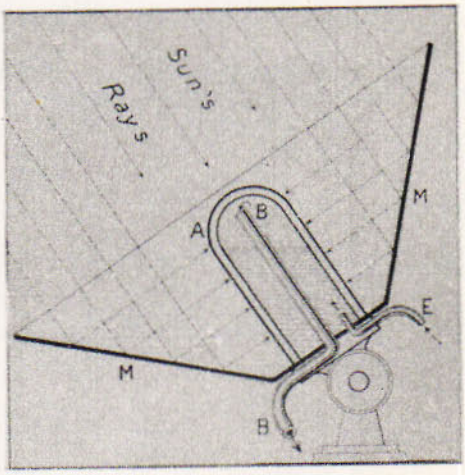


Fig. 533 — PRINCIPLE OF A.-B. MOUCHOT'S SOLAR GENERATOR. M, mirror; A, boiler; E, water inlet; B, steam outlet.

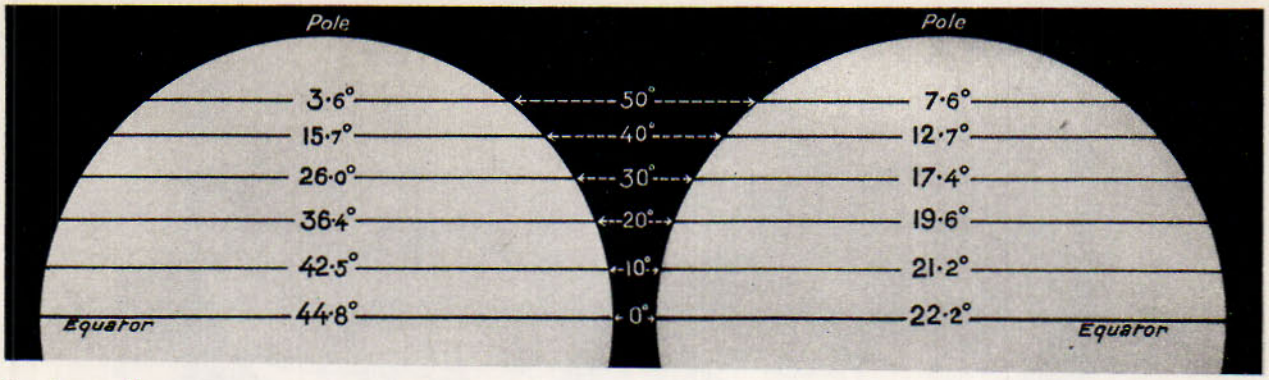


Fig. 534 — THEORETICAL TEMPERATURE DISTRIBUTION OVER A HOMOGENEOUS TERRESTRIAL SURFACE. Left: hemisphere of dry land; right: hemisphere of ocean. *(After Hann)*

of the 11-year cycle), and the causation of the vortical tunness from which the spots are derived.

Some advance in our knowledge of the Sun has nevertheless been made, and these speculations perhaps indicate the general direction which future research will take.

Correlations between solar and terrestrial phenomena. — The exact value of the intensity of the Sun's radiations was for a long time unknown. That its output of heat and light is gargantuan is obvious enough, but the Sun's total emission is not limited to these, and its additional radiations cannot be directly apprehended by any of our five senses, though they often have more or less marked indirect effects. Furthermore, to these electromagnetic radiations, visible and invisible, must be added the Sun's corpuscular emission, whose existence is indicated in numerous ways. Taken as a whole, the Sun's emission represents a prodigious output of energy.

From measurements of the quantity of solar heat received by the Earth (see p. 283 — solar constant) it can easily be calculated that every square centimetre of the Sun's surface continuously emits a flow of 1550 calories per second. One need only consider for a moment the number of square centimetres comprising the whole solar surface to gain some idea of the torrent of energy that the Sun pours out unceasingly.

But instead of expressing the value by abstract numbers, whose very magnitude renders them almost meaningless, let us try to convey some idea of the Sun's radiation by means of concrete analogies with familiar things: we can calculate, for example, that in order to produce the same quantity of heat as the Sun, 11 thousand million million tons of coal would have to be burnt every second — or 11 million times more than that produced by all the coal mines of the world in a year.

We shall see later (Chapter 15) how modern theory has identified the source and explained the continued production of this stupendous output of radiation. For the moment we shall concern ourselves with the effects that it has upon the life of our own planet.

Illumination and heating of a planet. — The intensity of the radiation falling on different regions of a planetary globe will always be very unequal. At the point on its surface directly opposite the Sun (i.e. the point at which the Sun is overhead) the radiation will strike the ground perpendicularly. As one moves further away from this point, approaching the edge of the illuminated hemisphere, the Sun's rays will arrive more and more obliquely, till finally, at sunset or sunrise, they merely graze the surface. These varying conditions of illumination and their consequences can easily be illustrated by a diagram. Suppose that RR (Fig. 527) is a beam of solar radiation of given diameter. Arriving perpendicularly, it illuminates a surface area XY. If somewhat oblique, it will illuminate the large area XY′; if still more oblique, approaching the horizontal, the still larger area XY″. It is obvious that the intensity of the radiation received at the surface is proportional to the elevation of the Sun above the horizon.

These differences — which regulate the general heating and lighting of the body of a planet by the Sun — can be directly observed, at any rate so far as illumination is concerned. A photograph of the Moon, for example, shows clearly that towards the Quarter the brightness of the sunlight falls off progressively as the terminator is approached. The same effect is equally clearly seen when Mercury or Venus is observed telescopically.

The importance of this effect is augmented when the planet is surrounded by an atmosphere. Referring to the diagram again, it will be seen that the radiation reaching the surface will be still further attenuated as a function of increasing obliqueness by absorption in the atmospheric layer, AA: the more oblique the incident ray, the greater the thickness of atmosphere it has to pass through. We have already considered the terrestrial effcts of this absorption, but it is worthwhile remembering that the rôle of the atmosphere is not only important but very variable, owing to its lack of homogeneity and the complex disturbances to which it is subject. Let us, however, ignore these details and concentrate on the main principles involved; by this means we shall be able to elucidate the general conditions prevailing on any particular planet.

The general distribution of heat and light is thus theoretically established: progressive decrease of intensity from a maximum at the sub-solar point to a minimum at the periphery of the sunlit hemisphere (ignoring the night hemisphere). Let us apply these conclusions to different planets.

The planets rotate more or less rapidly, their axes maintaining a definite orientation relative to the planes of their orbits. If the axis is normal to this plane, the poles will always by situated at the edge of the sunlit hemisphere — i.e. will always receive minimal heat and light from the Sun — while the equatorial region will experience the whole range of intensity from maximum to minimum. But there are two different cases to be considered: rotation in the same period as the revolution, and rotation more rapid than this.

In the first case — exemplified by Mercury, for instance — the same hemisphere is always turned towards the Sun, while the other lies in perpetual darkness. In these circumstances, the classical theoretical distribution, as described above, will apply: minimum insolation along the edge of a hemisphere from pole to pole, with a single point on the equator receiving maximum heat and light.

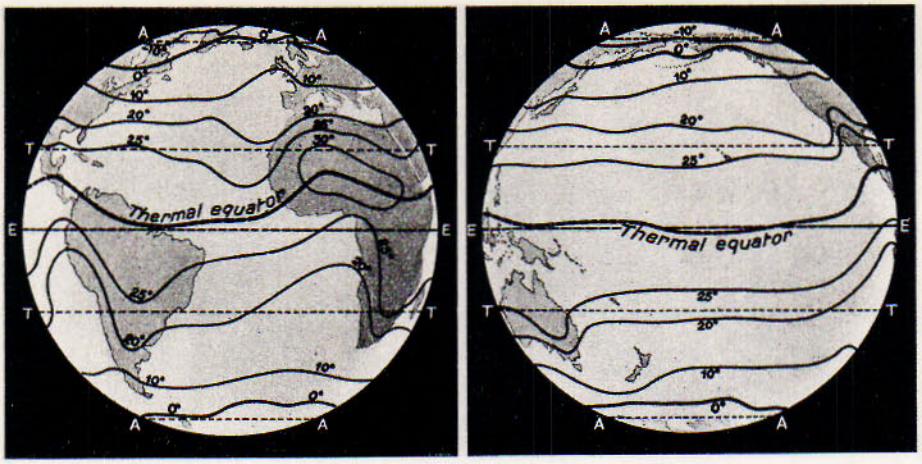


Fig. 535 — ACTUAL ISOTHERMS AT THE EARTH'S SURFACE

Accordingly, at every point on the surface of such a world a unique and unvarying set of conditions will prevail.

In the second case — that of rapid rotation — every point except the poles will receive a regularly varying amount of radiation, repeated each day and progressively exaggerated as we approach the equator; here the difference between maximum and minimum will attain its greatest value, the Sun passing from horizon to horizon through the zenith.

The conditions regulating the maximum and minimum temperatures, however, are not quite the same as those applicable to intensity of illumination. The heating effect of the Sun's radiation will be cumulative, so that maximum temperature will not be attained at midday (when the Sun is at its greatest altitude), but some time later, the amount of this lag being a function of the speed of rotation. On the Earth, the interval averages 2 hours. It is, in fact, a matter of common experience that the early morning is, generally speaking, cooler than the late afternoon. Owing to a similar cumulative effect, July — and not June, when the summer solstice occurs — is the hottest month of the northern year.

In short, daily variations of insolation will occur over the whole surface of a globe whose axis is perpendicular to the plane of its orbit; but for a given point they will remain always the same — in other words, there are no seasons.

Seasons modify the amplitude of the daily variations, in one direction or the other, and arise from the inclination of the axis of rotation. Let us recall the explanation already given of the mechanism of the terrestrial seasons, which produce considerable inequalities in the length of the day; these increase from the equator, where they are nil, to a maximum at the poles. Each latitude thus receives an unequally varying insolation from one season to another. These effects are exaggerated as the inclination of the axis increases.

It is unnecessary to emphasise the complications that are introduced if the planet has an atmosphere, whose complex changes are themselves caused primarily by the Sun's radiation, though they also depend on its relative density or tenuity. The speeds of rotation and revolution, and hence their respective durations, similarly play their part; identically situated points on the surfaces of two planets which differ in this respect will be exposed to different periods of insolation, which in addition will be of unequal intensity. It must also be borne in mind that the intensity of the received radiation is a function of the planet's distance from the Sun, decreasing with this distance according to the inverse square law. The intensity at the surface of a planet 10 times further from the Sun than the Earth (Saturn, for example) will be 100 times weaker.

All this has been treated in some detail since it helps us to realise the diversity of the conditions reigning on the different worlds of the Solar System. And we have only

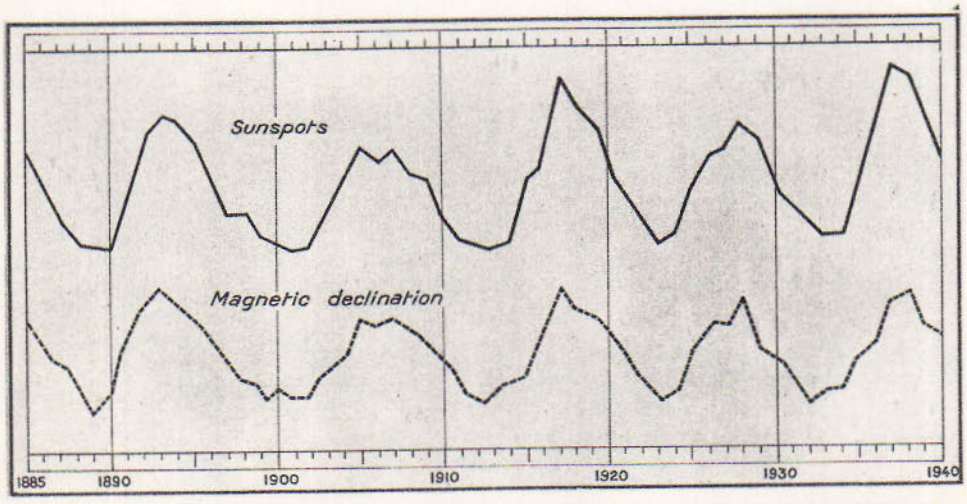


Fig. 536 — THE VARIATION OF SOLAR ACTIVITY COMPARED WITH THE MEAN AMPLITUDE OF THE DIURNAL VARIATION OFJMAGNETIC DECLINATION

Fig. 537 — THE AURORA BOREALIS IS ONE OF THE MOST BEAUTIFUL MANIFESTATIONS OF THE SUN'S ACTION UPON THE EARTH'S UPPER ATMOSPHERE. General appearance of the aurora of 6 February 1939, observed by L. Rudaux at Donville (France).

been considering the theoretical picture. When we attempt also to take into account the local conditions and peculiarities of each planet, the problem becomes too complex for precise conclusions to be arrived at: too many of the relevant factors are unknown — the differing effect of the atmosphere from one region to another, for instance, which depends on its composition, density and extent, as well as on its detailed features.

The solar constant. — The foregoing account has explained how, and under what circumstances, the globe of a planet receives the Sun's radiation.

We must now investigate in more detail the actual amount of solar energy that is received by the Earth — the one place where we can directly assess its effect. The importance of the rôle that this energy plays for mankind, when converted into heat, needs no stressing.

Such measurements must allow for the effect of the atmosphere, whose upper regions diffuse and reflect a certain fraction of the Sun's radiation back into space. That which penetrates the main mass of the atmosphere is ultimately, after a somewhat complex process of degradation, converted into long-wave energy for the maintenance of the Earth's heat: the measure of this energy is termed the solar constant, and is expressed in heat units.

The mean value of many independent determinations of the solar constant — defined as the quantity of heat received per minute by 1 sq. cm. of the Earth's surface exposed to the vertical rays of the Sun — is 2.0 calories, a correction having been made for atmospheric absorption. On account of this absorption we receive at most three-quarters of the energy incident on the outer limits of the atmosphere.

These determinations are of the greatest delicacy, for their accuracy demands that they be freed from the effects of numerous accidental factors — the altitude of the observing station, the season, the hour of the day — all this in addition to the somewhat inadequately determined absorption factor. Despite all these difficulties, sufficient accuracy has already been attained to show that the term solar 'constant' was perhaps ill-chosen, for the discussion of all the available determinations appears to indicate that its value varies. Such, at least, is the conclusion drawn by Abbot and his colleagues from their classic investigations.

Variations are certainly recorded, but it is difficult to be certain of their objective significance, owing to the ever-changing conditions of the atmosphere. Some of these variations are fairly sudden; others are of longer period and are more securely established since they are apparently linked with the curve of the Sun's cycle of activity. Their magnitude is of the order of 1% of the value of the constant.

Turning now to the practical effects of this radiation, it can be calculated that assuming mean atmospheric transparency each square centimetre of the Earth's equatorial region receives 250,000 calories annually. Altogether, this represents enough heat to melt a layer of ice 100 feet thick and surrounding the entire equator; or, alternatively, to convert a sheet of water 13 feet deep into steam. The figures for the rate of emission at the Sun's surface have already been quoted.

Attempts to harness and utilise the Sun's heat — for example, by concentrating the solar radiation at the focus of appropriately arranged convex mirrors — therefore suffer from no lack of raw material. Such attempts are at least as old as Archimedes' burning mirrors, references to which are found in many of the writers of antiquity.

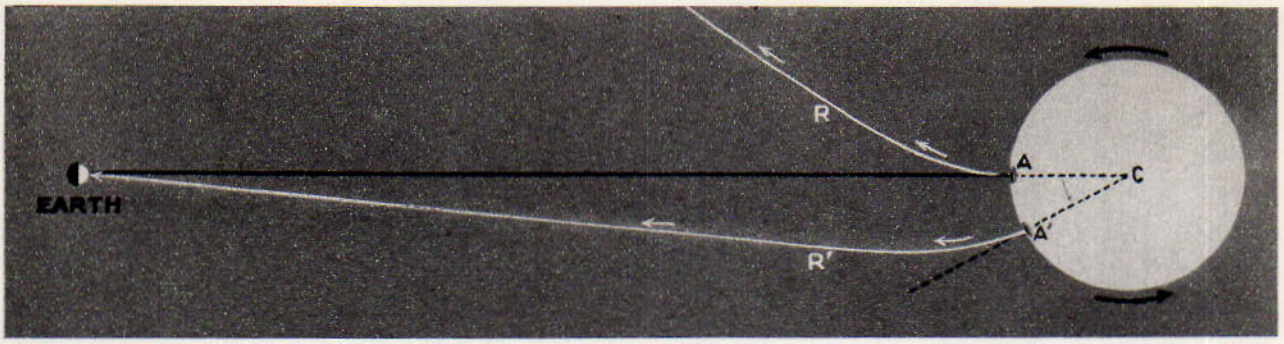


Fig. 538 — EXPLANATION OF THE TIME-LAG BETWEEN MAGNETIC DISTURBANCES AND THE MERIDIAN PASSAGE OF THE SPOTS THAT CAUSE THEM

No doubt the reader himself has performed the simple experiment of concentrating the Sun's heat with an ordinary lens, and igniting a cigarette, some tinder, or a dry leaf held at its focus. There could hardly be a cheaper method of producing fire.

The utilisation of solar heat for industrial purposes was first advocated by Mouchot, who in 1861 envisaged a system consisting of a boiler located at the focus of a large conical or parabolic mirror which concentrated the Sun's radiation on it to drive a small steam engine. Later installations, constructed on a larger scale and employing more modern developments of the same principle, have succeeded in producing an appreciable output.

Unfortunately it has to be admitted that such plants have not yet been sufficiently perfected to allow of the effective utilisation of this free and inexhaustible source of energy. The Sun must not only shine: it must do so regularly and constantly. It is thus a waste of time to attempt the harnessing of the Sun's energy in more than a few regions of the world — those where it reaches the greatest altitude above the horizon (i. e. in the tropics and sub-tropics), and preferably, also, where the local meteorology provides long periods of clear skies.

In space, though, it is a different matter, and the solar batteries which have already been employed to power the instruments aboard earth satellites presage the future possibility of utilising solar radiation undimmed by any atmospheric effects whatever. This, however, is another aspect of the question into which we cannot here enter fully.

The Earth's climate. — The Earth, in travelling around the Sun, receives the latter's radiation in varying intensities and distribution, and it is to these variations that we owe our regular seasonal changes of climate. It is easy enough to give an elementary account of these large-scale variations, their progress and their relative importance. What is less easy is to account for the many small-scale variations which are so obvious a feature of the terrestrial weather, and of which, rightly or wrongly, we complain so bitterly — either it rains too much or not enough, the temperature is now excessively high, now intolerably low; on top of which there are the storms, the hurricanes, the cyclones. Nor is there any generality about these short-term vicissitudes, for simultaneously other regions will be experiencing the opposite conditions. Since not even the textbooks of meteorology can provide a thorough treatment of all these questions, let us confine ourselves to those disturbances which derive from the large-scale movements of the atmosphere. In this sphere the heating effect of the Sun is the prime mover, though their development is modified by such geographical factors as the distribution of the oceanic and continental areas and the nature and contours of the terrain.

The consequences of these surface irregularities are strikingly shown if we compare our actual climate with that which the Earth would have if it were a perfectly regular and homogeneous spheroid. To illustrate this, let us consider a single climatic factor — the mean annual temperature, which decreases with increasing latitude on either side of the equator.

This decrease, whose rate can be described by theory, would be regular if the Earth's surface were uniform — either entirely land or entirely sea. In the former case (as can be seen from Fig. 534) the isotherms* would show a greater difference between extremes than in the latter. The reason for this is simple: land heats up more quickly, but water conserves its heat better.

Differences of this sort are produced by the lack of uniformity of the Earth's surface. They modify, without of course annulling, the climatic effects of differences of latitude. In Fig. 534, the isotherms are shown parallel to the equator and therefore coincident with the parallels of latitude. The actual distribution of the isotherms, however, is not symmetrical and coincident with the parallels of latitude (see Fig. 535) since it is modified by

* A line on which every point has the same temperature.

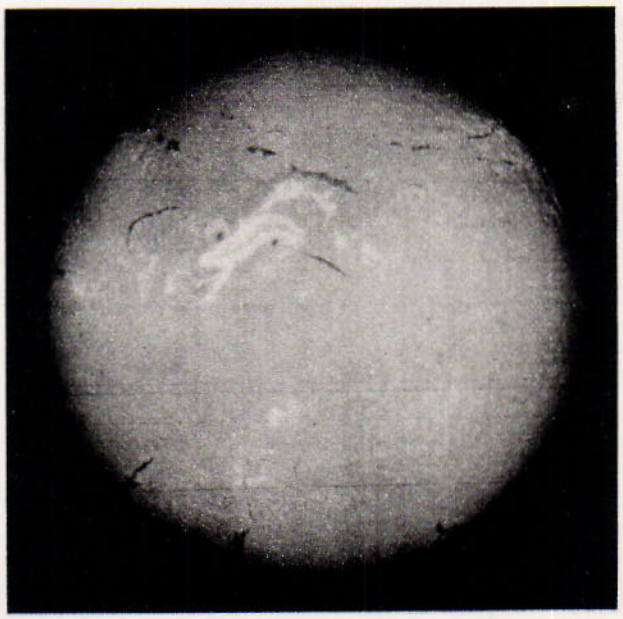


Fig. 539 — Hα SPECTROHELIOGRAM SHOWING THE GREAT FLARE OF 25 JULY 1946, 17th 32m
(Photo, Meudon Observatory)

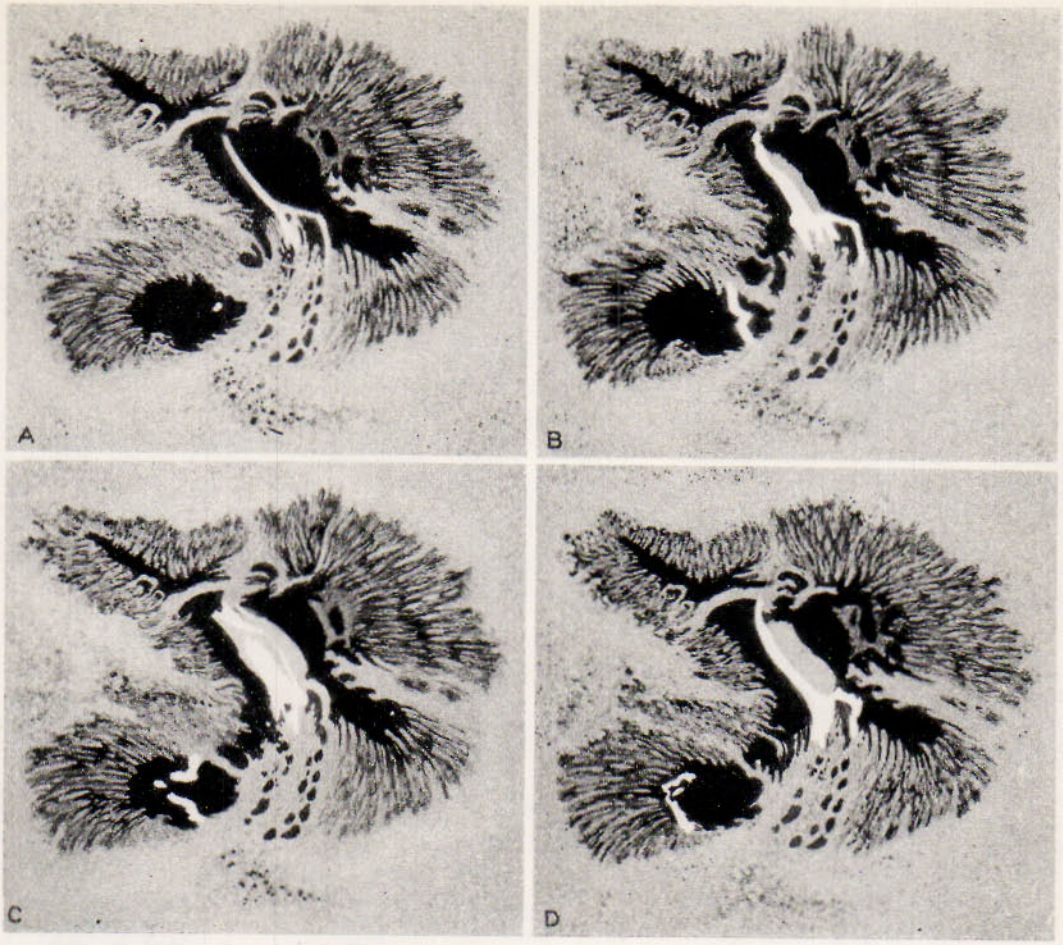


Fig. 540 — FLARE OBSERVED IN INTEGRATED LIGHT BY L. RUDAUX, 15 JULY 1892

the interplay of antagonistic influences — the oceans and the continents — and also by the individual characteristics of the land masses. Marine and atmospheric currents, warm or cold, also introduce anomalies in either direction, between different points in the same latitude.

Finally, geographical factors and the large-scale movements of the atmosphere produce very marked differences between regions which, so far as astronomical causation goes, should be climatically indistinguishable. Thus Norway, for example, whose mean annual temperature is above freezing point, lies in the same latitude as parts of Canada and Siberia which have a much lower temperature and a truly arctic climate.

The implications for the organised life on our planet of our infinitely varying and variable climate lie, for the most part, outside the scope of this book. Nevertheless, we may refer briefly to certain facts concerning the Earth's vegetation which are directly attributable to the fluctuating light and heat received from the Sun. Plants, unlike animals, cannot migrate in order to escape from intolerable environmental conditions: they must adapt or die. Some species, of course, are more adaptable than others, but no matter what the species, there are limits to what it can survive. These limitations of the permissible habitat are imposed by altitude and latitude, whose effects are in many respects similar: high altitudes impose rigorous conditions similar to those of the arctic regions.

In many different ways, vegetation is an eloquent witness to the climate in different regions; it must have light (without which the chlorophyll reaction cannot take place), warmth, and moisture. Thus the world-wide distribution of vegetation is primarily determined by large-scale climatic zones: it flourishes abundantly in equatorial regions, becomes more and more thinly distributed throughout the temperate regions, and finally dies out altogether in high latitudes. This simplified picture ignores the more or less extensive desert regions, whose adverse conditions result from large-scale movements of the atmosphere causing a general deficiency of rainfall or even its complete and prolonged absence; thus regions with extremely sparse vegetation, adapted to an unfavourable environment, are to be found in the same latitudes as areas which are covered with the most luxuriant virgin forests.

The effects of climatic extremes, whose causation can be regarded as purely 'astronomical', are very clearly marked. Towards the edges of the vegetated region (whose extent is determined by the temperature), trees and then bushes die out first, finally leaving only plants, which tend to cling closely to the ground, over which they spread as though to derive the greatest benefit from its warmth and to resist the ravages of the intemperate climate. Mountaineers have likewise commented upon the extremely vivid colours of the flowers at great altitudes, which may be attributed to the intensity of the light owing to the increased transparency of the atmosphere at these heights.

The daily duration of sunlight — a direct consequence of latitude and season — also plays a very important rôle. At certain seasons it can compensate for the low mean temperature in regions where the Sun, at other times of the year, remains below the horizon for days or weeks at

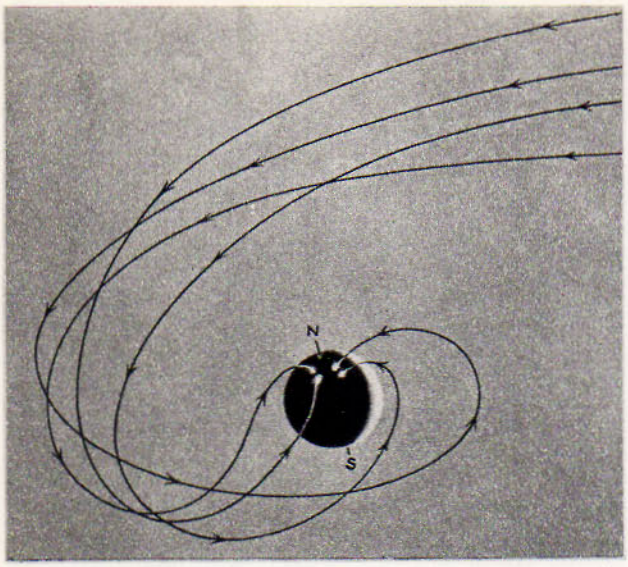


Fig. 541 — PATHS WITHIN THE TERRESTRIAL MAGNETIC FIELD OF THE STREAMS OF ELECTRONS FROM THE SUN WHICH ARE RESPONSIBLE POLAR AURORAE *(After Stormer)*

a time. An instance of this was noted by the celebrated American explorer Peary, who in northern Greenland came across a veritable oasis, a riot of plants and flowers (the nature and exposure of the soil also being suitable, of course); for although it was situated midway between the Arctic Circle and the pole, the summer Sun in these high latitudes does not set for 16 weeks together.

In general outline, then, the establishment and proliferation of vegetation emphasises the importance of the climatic differences which we have just been discussing. But it is not only through the agency of its light and heat that the Sun affects the Earth. It also influences our world in a more subtle fashion, whose character was for a long time not properly understood.

Effects of the Sun's corpuscular, ultra-violet and long-wave radiation. — We have already spoken of these effects in connection with the Earth's magnetism and kindred matters. It has been established that, in this field, the Earth is subject to disturbances which originate in the Sun and are more or less intimately connected with observable solar events.

The subject is of importance to the geophysicist as well as to the astronomer, for some of these effects have practical repercussions which affect us personally in our daily lives. The phenomena are very complicated, however, and are still far from being completely understood.

The influence of the Sun on the Earth in this respect is most obviously shown by the very close correspondence between the periodic variation of the terrestrial magnetic field and the solar cycle as indicated by spot numbers: even a quick glance at Fig. 536 will tell more than a page of description.

This remarkable correspondence concerns the mean amplitude of the diurnal variation in magnetic declination. But in addition to these continuous daily variations, sudden disturbances also occur from time to time — the so-called magnetic storms, which are usually accompanied by auroral displays; it will be remembered that the latter are most frequently observed in high latitudes, but that they are nevertheless not rare in the latitudes of Europe.

Magnetic storms and aurorae are in general more frequent at and around sunspot maxima. In addition, they very frequently occur when a large spot or group of spots has just crossed the Sun's central meridian. It is also to be noted that the severity of the terrestrial disturbance bears no relation to the area of the spot: many spots, indeed, produce no magnetic disturbances whatsoever.

These facts are interpreted as follows: radiation emitted by the spots, or by the eruptive developments associated with them, is in part composed of a stream of electrons which is projected into space in a specific direction, like water from a hosepipe; this jet ultimately encounters the Earth. Because of the Sun's rotation, the direction in which the jet is emitted does not remain perpendicular to the solar surface (see Fig. 538), but deviates progressively in the direction R, opposite to that of the rotation. Thus a jet does not reach the Earth when the active region, A, is directly facing us, but when it is in position A′, some little time after its meridian passage; this lag is measured in terms of the angle through which the Sun has turned during the interval.

By means of observations of this sort, it has been established that the corpuscular radiation requires about a day to cross the void separating the Sun and the Earth. More precisely, since some magnetic storms are of long duration, their onset follows the causative solar event after an average interval of from 25 to 30 hours, while their maximum intensity is attained after about 45 hours.

These magnetic disturbances are the signal that the streams of electrified particles have arrived at, or in the neighbourhood of, the Earth, where their motions set up a secondary magnetic field in accordance with the classical laws of electromagnetism.

In this explanation we have been considering the case of an electron stream which is so oriented as to impinge directly on the Earth. A magnetic storm is always to be linked with some disturbance in the surface layers of the Sun which has recently crossed the central meridian; the reverse, however, is not true, important spots often being observed to cross the Sun's disc without any magnetic storm occurring on Earth. Indeed, the position of the disturbance on the solar surface may be such that its emissions would miss the Earth altogether, passing either above or below it. Moreover, the Earth is so small and the Sun so distant that from a solar point of view it is no more than a point in space. For an electron stream to hit the Earth is rather like scoring a bull's-eye on a target that is all but invisible.

Although violent magnetic storms are caused by the arrival of streams of relatively slow-moving electrons (several thousands of miles per second), other disturbances seem to be caused by radiation travelling at the speed of light. This distinction was hinted at the first time a visually observed and sharply defined solar event was shown to be intimately connected with an equally sharply defined magnetic disturbance. This classic observation was made by Carrington and Hodgson on 1 September 1855, when they observed the sudden outburst of some very brilliant flecks of light in the midst of a fine spot group; these were subject to very rapid displacements before they disappeared again, about 5 minutes later. At precisely the same moment as they observed this outbreak, the magnetic recording apparatus at Kew registered a weak and short-lived disturbance; then, about 18 hours later, a very violent magnetic storm broke out, and continued for several days.

A similar and equally instructive observation was made on 15 July 1892 (see Fig. 540). The short-lived and rapidly moving bright patches — which in a space of 20 minutes modified the whole appearance of the spot — were of blinding intensity despite the dark filter which is usually employed when observing the Sun. But, more important, terrestrial repercussions accompanied and

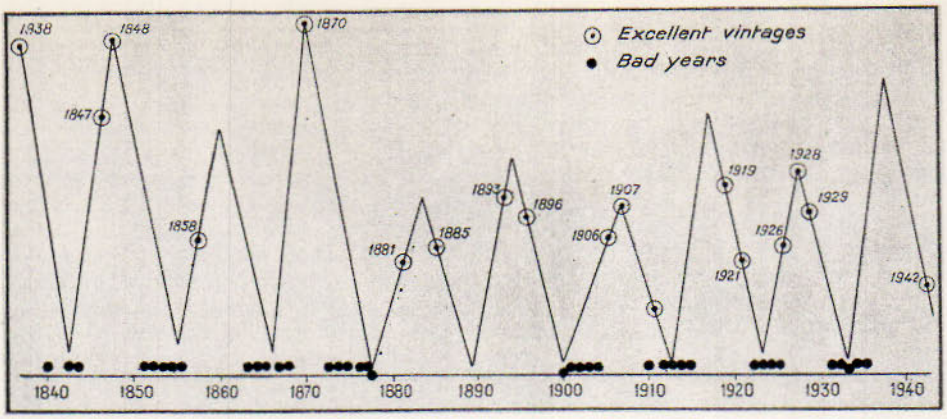


Fig. 542 — SUGGESTED CORRELATION BETWEEN THE QUALITY OF BURGUNDIAN WINES AND SOLAR ACTIVITY *(After G. Bidault de l'Isle)*

followed this observations just as they did that of Carrington and Hodgson: at the precise moment of the visual observation, a slight disturbance of the magnetic elements (declination, horizontal component and vertical component) was recorded; then, 19½ hours later, came the sudden outbreak of a violent magnetic storm, which persisted for 32 hours and caused the interruption of telegraphic communications in Spain.

The solar features which thus testify so eloquently to their action upon the Earth are of relatively frequent occurrence. It is only rarely that they are visible with a telescope, as in the two cases just described. Owing to their suddenness and short durations an observer has but small chance of observing the Sun at precisely the right moment — the more so since they are normally produced under conditions which do not favour their visual observation. Modern observational techniques have fortunately multiplied the occasions on which they may be observed. They are known as flares, and are particularly accessible to investigation by the spectroheliograph or the visual form of the same instrument, the spectrohelioscope. A continuous patrol of the chromosphere (subject only to interruptions from the weather) is nowadays kept by a world-wide network of these instruments.

Despite the progress that has been made, there still remains a great deal to be explained — for example, terrestrial disturbances for which no corresponding causative event can be detected on the visible hemisphere of the Sun. This phenomenon — which would appear to be inexplicable in terms of electronic streams directed towards the Earth — is explained, according to Dauvillier's theory, by the very great curvatures imposed on these streams after they have left the chromosphere. The conditions shown in Fig. 538 should, in fact, be exaggerated still further; thus the active zone producing the stream may not even be situated on the visible hemisphere at the time it reaches the Earth. Curvatures of this magnitude may perhaps be induced in electronic streams by the Sun's general magnetic field.

But let us revert to the more generally agreed facts.

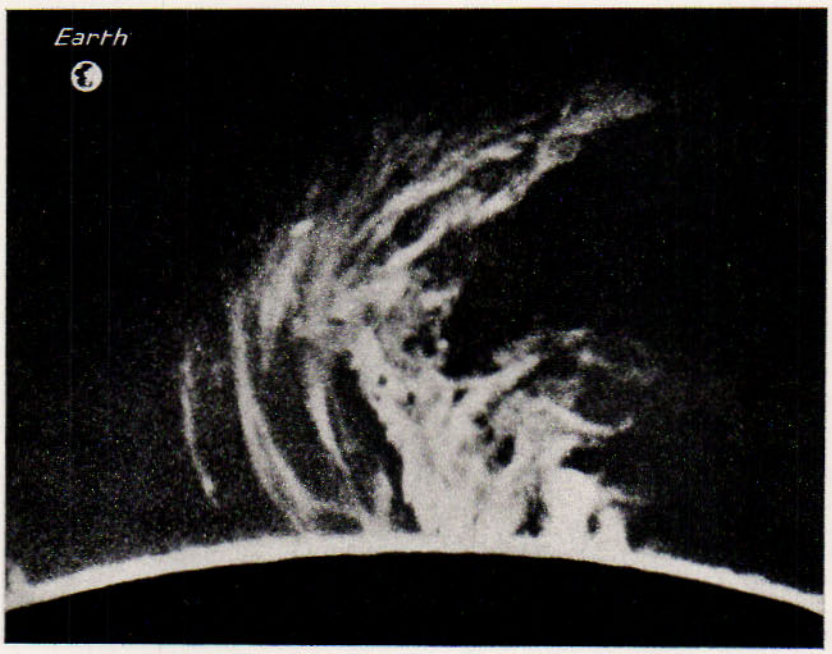


Fig. 543 — THE GREAT ERUPTIVE PROMINENCE OF 9 JULY 1917, WHICH REACHED A HEIGHT 140,000 MILES, PHOTOGRAPHED WITH THE K LINE OF IONISED CALCIUM. The small disc in the top left-hand corner shows the comparative size of the Earth. *(Photo, Mount Wilson Observatory)*

Although the flares are violent, brief, and subject to recurrent outbursts, the regions producing them may maintain a high level of activity for considerable periods. Some magnetic disturbances have, indeed, been repeated at intervals of 27 days — this, being the Sun's mean synodic period, is the interval between successive arrivals of a given area at the same position on the disc relative to the Earth.

The different intervals required by the Sun's radiation to reach the Earth suggest that it is not all of the same character: some make the journey at the velocity of light, some very much more slowly. Let us examine these different cases. We know that in addition to electromagnetic radiation, which is emitted impartially in all directions, the radiation producing terrestrial magnetic effects also contains another component which assumes the form of relatively narrow jets or streams. The terrestrial effects are, first, small disturbances of the Earth's magnetic field, and secondly, directly observable phenomena such as the fading of short-wave radio transmissions and the strengthening of those on long wavelengths.

The explanation of the mechanism of these disturbances involves the upper levels of the Earth's atmosphere — the ionosphere — and the electrified layers at different altitudes that comprise it. The long-distance propagation of radio waves depends upon their reflection from, or refraction in, these ionised layers, and its interruption is caused by the formation of large quantities of new ions which modify the normal characteristics of the ionosphere.

These ionisation effects in the upper atmosphere are attributed to radiations in the extreme ultra-violet, which are emitted in considerable intensity by solar flares. It should be added that another section of the Sun's ultra-violet spectrum is responsible for the ozone layer and also maintains the permanent ionisation of the upper atmosphere. This ionisation varies as a result of varying illumination between day and night, winter and summer, and of the probable fluctuations of the overall intensity of the Sun's ultra-violet emission from one year to the next.

On the other hand, the 'directional' emissions which are responsible for violent magnetic storms and aurorae, and which take 20 hours or longer to reach the Earth's orbit, are corpuscular — more precisely, electronic.

These beams of negatively charged particles are, in other words, identical with the cathode rays of our laboratory vacuum tubes. According to Deslandres and Birkeland, they also provide the best explanation of the formation of aurorae; the investigations of Störmer, the Norwegian physicist who has made a special study of these phenomena, lead to the same conclusion.

When electrically charged particles encounter a sufficiently strong magnetic field, their normally linear paths are distorted into curves, and these reproduce more or less closely the lines of force of the field; they tend to concentrate in those regions where the intensity of the field is greatest and the lines of force are consequently most closely crowded together. Birkeland has shown experimentally that cathode rays bombarding a magnet reproduce the forms of the auroral ray. And Störmer, calculating the paths that solar electrons would follow when they reach the Earth's vicinity and come under the influence of its magnetic field, has shown that they would congregate mainly in the night-side hemisphere, and in two high-latitude regions (one north and one south) centred on the magnetic poles. These zones correspond exactly with those in which the greatest numbers of aurorae are observed. There is now no doubt that the light of these beautiful objects results from the excitation produced by the solar electrons on their arrival in the upper atmosphere; the process involved is very strictly analogous with that which occurs in domestic strip-lighting tubes.

The various types of radiation so far described do not, however, represent the sum total of the Sun's emission. A previously unsuspected radiation was discovered during the last war by radar operators whose sets inexplicably went haywire when directed towards the Sun: it also, is associated with sunspots and is electromagnetic in nature, but of very long wavelength. It in fact belongs to the domain of short radio waves, with wavelengths of several metres; the receiving sets transform it into a whistling sound or a background noise which starts suddenly, or at least is suddenly strengthened, at the moment when a fade-out occurs on the short waveband at about 10 metres. Little is yet known of the origin of this long-wave electromagnetic emission, whose study is still in its infancy. It is generally supposed that its production is due to the vortical motions of electrons in, or rather above, sunspots; there is, as we have seen, plenty of evidence of such motions.

The solar cycle and natural terrestrial phenomena. — The close correlation between variations of terrestrial magnetism and the 11-year cycle encourages us to look for similar correlations among other natural phenomena.

We have seen how the flood of solar radiation in which the Earth is bathed produces complex meteorological changes which, in their turn, directly affect the organic life of the planet.

It must be repeated that we are ignorant of the fundamental cause of the 11-yearly upsurge of solar activity. The development of such features as the spots is no more than a spectacular manifestation of this activity. But this more or less marked periodic effervescence does suggest that possibly the Sun's generation of heat and light is not constant.

This indeed would seem to be proved by the delicate determinations of the solar constant, which reveal a fluctuation whose maxima coincide with those of the spot cycle. At first sight this might seem difficult to reconcile with the relatively low temperatures of the spots. No real paradox is involved, however, since everything points to the fact that the spots are to be regarded as symptoms of a general stoking-up of the solar furnace.

It has been calculated from the amplitude of the variation in the solar constant that the mean temperature should vary by 1.5° C between spot maxima and minima. This is a small difference, and one for which it is difficult to produce evidence, since local observations are entangled in other variables — geographical and meteorological primarily — which are large enough to swamp it.

One should rather look for a general effect in the large-scale seasonal phenomena, such as the length or warmth of the summers or winters, or the abundance or lack of precipitation. The problem is, admittedly, rendered excessively complicated by reason of the number of factors involved, whose weight has not yet been accurately estimated. It is not altogether surprising that many of its aspects have not been elucidated.

To obtain some idea of these difficulties, let us suppose that the Earth receives an increased amount of heat. The greater part of the Earth's surface is covered by oceans, so that the first result will be an increase of evaporation; this in turn (by a well-known law of physics) will produce a lowering of the temperature. What will be the result of these two conflicting tendencies? In one case, one may counteract the other; in another case, one may prove predominant; the final result will always depend upon a multitude of local conditions. Apparently contradictory

results, therefore, cannot but be expected. According to the by no means recent investigations of Koppen, for instance, the mean temperature at the equator should be slightly lower when the Sun's activity is maximal than at other times. Yet in other regions the exact contrary has been found; and in others again, a correspondence between the two variations.

Rainfall seems to be capable of furnishing a more precise criterion. If there is an increase in the heat received by the Earth during solar maxima, the atmosphere, as we have just seen, will become richer in water vapour; hence increased rainfall may in general be expected, though this will to some extent lag behind the solar phase. This effect will only be detectable if vast areas of the Earth's surface are considered: the quantity of rain falling on such an area will (assuming that it drains away) be reflected in the level of any natural reservoirs for which the area is the catchment. This appears to have been demonstrated in the case of the Caspian Sea and of the great African lakes, Albert and Victoria.

A remarkable 30- to 35-year periodicity in the meteorological phenomena of western Europe may possibly be correlated with a combination of three solar cycles. This period, discovered by Professor Brückner, is itself divided into two phases, each with a mean duration of from 15 to 17 years: a series of cold, damp years followed by one of warm, dry years.

T. Moreux, again, has suggested the existence of a connection between solar maxima and the increased speed or extent of the melting of the polar ice. Antoniadi believed that it is possible to establish a similar correlation in the case of Mars.

In another direction, since the characteristics and behaviour of organic life are modified by large-scale seasonal changes, attempts have been made to discover parallels in this sphere. Thus an interesting series of observations made by Camille Flammarion showed that, in our region of the world, the flowering of chestnuts and lilacs, and the return of certain migrants such as swallows, cuckoos and nightingales, seem to be influenced to a marked degree by the fluctuations of the Sun's activity.

It should be added that simple statistical data appear to reveal a dependence on the solar cycle of widescale agricultural production, such as that of wheat and wine. In the latter connection, a recent enquiry undertaken by G. Bidault de l'Isle should particularly be mentioned. His aim was to establish a connection between the solar cycle and the quality of Burgundian wines. An exact parallelism between the two was not to be expected, owing to purely local meteorological factors, but he succeeded in producing evidence that the good years show a preference for epochs of maximum solar activity, whilst during minima the vintages were inferior. These interesting conclusions at least corroborate those already quoted (Fig. 542).

In another sphere again, medical research has apparently uncovered an occasional increase of certain mental ailments, suicides, etc., at periods of sunspot maxima.

Among physical phenomena which have a well-established connection with the Sun's activity may be mentioned — though its explanation is still unknown — the brightness of the eclipsed Moon; this brightness (as we saw in Chapter 6) is conditioned by the comparative overall transparency of the Earth's atmosphere, and particularly of its upper regions. This relation has already been described in the chapter devoted to eclipses, and all that need be said here is that the characteristic serrated form of the variation makes it very difficult to identify its cause; it bears no relation to the regular fluctuation of the total area of the Sun's surface covered by spots, for example, and is suggestive rather of the latitude variation of the spots during the course of the cycle. Nevertheless, it is difficult to see how this could affect the transparency of our atmosphere.

Many other correlations could doubtless be established, for the Earth's dependence on the Sun is as comprehensive as it is complex; a whole book could, indeed, be devoted to its discussion. But the examples that have been given provide sufficiently convincing proof that all the phenomena of Nature are more or less directly influenced by the rhythm of the Sun's activity. This is not to say that the problem is solved, however. It must in fairness be admitted that in many cases further confirmation or increased precision are required. Thus much work still remains to be done: first, in defining more precisely the progress of the solar cycle, with its sometimes very great irregularities; and then in discovering the nature of the interactions between our world and the Sun, with its prodigious output of radiation of all kinds; and finally, in identifying their meteorological and other consequences. Increased understanting of these interactions would not only be of interest to theoreticians: its practical applications could also be immense.

Fig. 544 — A BEACH PHOTOGRAPHED AT LOW TIDE BY THE LIGHT OF THE NIGHT SKY

CHAPTER II

IN SPACE

The study of the Universe is not limited to the examination of the worlds that populate it, the unravelling of their motions and distribution, and the determination of their chemical and physical natures. We must also consider the phenomena of space itself — the vast ocean in which these bodies move and have their being.

The nature of this domain was long misunderstood, although some of its phenomena are visible to all. It is, nevertheless, understandable that these should have been denied proper attention, since their appearance is by no means striking or spectacular. Indeed, the principal one may easily be taken for granted and thus escape notice altogether, so obvious is it: the familiar appearance of the starlit sky.

Darkness and night vision. — After sunset, when the Sun has ceased to illuminate even the highest levels of the atmosphere above us and night has really fallen, the darkness should be complete, and black as pitch. And the testimony of our ordinary senses suggests that it is; the truth, however, is somewhat different.

When night has really fallen — it being taken for granted that there is no Moon — the darkness is in fact always something less than complete. Even a superficial examination of the night sky will show that the celestial background always retains a certain luminosity. This luminosity is demonstrated by the fact that buildings, trees and the like can be seen sharply silhouetted against it. Furthermore, it produces a faint general illumination of the scene which we shall discuss in a moment.

Even though this fact is, generally speaking, known to all, it is far from being commonly appreciated at its true value — witness many colloquial descriptions of night such as 'black as ink', or even 'pitch dark'. Misleading though such expressions may be, they are nevertheless partly justified by our normal conditions of existence, which do not fit us for night vision and whose effects take some time to wear off. Thus the great contrast between our artificially lit rooms, street lights, etc., and the dim illumination of the night sky militates against our proper appreciation of the latter. It is in fact only in the depths of the country that we encounter conditions which allow us to judge truly the degree of darkness of the night; furthermore, the retina takes some time to accustom itself to new conditions, and complete dark adaptation after exposure in an electrically lit room takes at least 15 minutes. Once our dark adaptation is complete, we will discover that there is always enough light for us to find our way about.

It is of interest to determine more precisely the limitations of effective night vision. Experiments have shown that the light of a 'dark' night is sufficient for printed

letters half an inch high to be read, and even the hands of a wrist-watch (if not too fine) can be well enough distinguished against the tiny black marks of the figures for the time to be told. The possibility of satisfactory vision under these conditions is nevertheless somewhat restricted, since visual acuity falls off very rapidly with increasing distance. Thus an object which is clearly visible at close range, quickly becomes ill-defined and finally invisible — depending, of course, on its size and the lightness or darkness of its tone — as we walk away from it. For example: a nearby telegraph pole is clearly seen in silhouette against the sky, but can be made out only with difficulty at about 50 yards, and vanishes altogether at a slightly greater distance. The individual branches of a tree are obvious enough when seen from near at hand, but a little further off are no longer seen separately, merging into an indistinct mass — an effect which becomes more pronounced with distance; the point at which this mass fades from sight depends on its size and density.

It is the same with light-toned objects or details. The individual units of a pile of stones or a wall, which at a distance of a few yards can be counted, merge at a slightly greater distance into an undifferentiated greyish surface.

The irregularities in an untarred, dirt or sand road — such as stones and potholes — can be avoided with ease when they are at our feet, and retain their outlines for 100 yards or so, according to their size. But it is, naturally, white objects which are most clearly visible, and remain visible at much greater distances than dark-toned objects of the same size. A man dressed in white, standing against a dark background, can be distinguished at a distance of about 80 yards. The war-time use of white paint for notices of all sorts, during the blackout, will be remembered.

Thus during the night-time we can in fact discern quite a lot of our immediate surroundings. This is not only true of completely cloudless nights, for though these naturally provide the most favourable conditions the night light nevertheless persists to some extent even when all the stars are obscured.

One final remark in this connection: when we say of a night that it is 'pitch black', we are referring to the degree of its darkness rather than to its colour. The general bluish tint of 'darkness' is commonly recognised, as is evidenced by the name 'midnight blue' given to a certain colour of cloth, and by many colloquialisms.

All this would appear to be firmly established, but it is not precise. Research into the problem of the brightness of the night sky has in fact shown that it is proportionately richer in the red and yellow wavelengths than the day sky. In other words, the night is reddish rather than blue. In giving us the opposite impression, our eyes are victims of the illusion known as the Purkinje effect, which has already been described in connection with the quality of moonlight (p. 125). Another reason is the tendency of the human eye to register complementary colours, so that by comparison with artificial lights, which are predominantly yellowish, the darkness appears to be bluer than it actually is.

What is the precise intensity of this nocturnal light?

Intensity of the night light. — A considerable amount of work has been devoted to answering this question, and it has been established that a horizontal surface exposed to the light of the night sky in absolutely cloudless weather receives as much light as it would from a candle at a height of about 180 feet above it. This figure assumes that the experiment is carried out in an open situation, from which the whole expanse of the sky down to the horizon is visible; for an enclosed site the intensity

Fig. 545 — THE DARK SILHOUETTE OF THE HOUSE EMPHASISES THE GENERAL BRIGHTNESS OF THE NIGHT SKY

of the illumination is decreased in proportion to the area of the sky that is hidden. Thus the surface of a street, or of a road bordered by tress, appears noticeably less bright than one in the open country. Town dwellers, when deprived of artificial lights, are therefore worse off than country folk.

Anyone unable to appreciate the night light for himself might imagine that it is non-existent, if he had only the figure quoted above to go on. And, indeed, all that we have just said concerning the brightness of the night sky demonstrates the marvellous powers of adaptation of the human retina, which can register and utilise so feeble an illumination.

Owing to the cumulative character of the photographic process, the camera is better fitted than the eye for recording very weak intensities of illumination — though it is not, of course, just a matter of taking snaps with a common-or-garden box camera. To obtain night photographs, one must use a very fast lens (i. e. one working at a large relative aperture, such as f/2 or f/1); even so, long exposures will be required — an hour or more, depending on the character of the landscape that is being photographed.

Photographic images possess another advantage over visual images. Whereas objects lying far from the centre of the visual field are not sharply seen, a well-corrected photographic objective gives perfectly clear images to the edge of the field, and reveals details which are indiscernible to the naked eye by reason of their fineness or their lack of contrast.

Though they require longer exposures, photographs taken at night have some of the same quality as those taken in daylight when the sky is overcast, for in both cases the light is diffused, so that objects are uniformly illuminated and cast no shadows.

Thus for the camera the word 'darkness' (applied to the night) is meaningless, and for the human eye it is purely relative.

Origin of the night light. — The brightness of the night sky may always have been recognised, but it is only comparatively recently that it has been systematically investigated in an effort to determine both its intensity

Fig. 546 — SNOWSCAPE PHOTOGRAPHED BY THE LIGHT OF THE NIGHT SKY

and its origin. These investigations have proved that its explanation is considerably less simple than was originally thought.

The appearance of the sky at night quite naturally suggests the idea of a 'dusky light falling from the stars', in Victor Hugo's phrase. Now although, obviously, these myriad stellar points must contribute to the night light, we shall learn shortly that they cannot be solely responsible for it. Once the intensity of this light had been accurately determined, it became evident that the total light of the stars is insufficient to account for it. The contribution of other phenomena, which in their turn demanded identification and explanation, had therefore to be invoked. So far as is known at present, these are to be found, first, in the depths of space, and secondly, on our own doorstep, as it were, in the upper levels of the Earth's atmosphere.

The problem thus posed turned out to be extremely complex, since the individual contributions of the various phenomena that are involved in the production of the night light might well be combined in variable proportions; furthermore, their respective contributions were for a long time impossible to disentangle, and it is only as a result of the work of the past 35 years that we can now identify them correctly.

However that may be, these investigations have led to some noteworthy results, perhaps the most remarkable being the abolition of the old idea of 'empty' space.

Starlight. — About 3,000 stars are visible to the naked eye at any one time — i. e. in one hemisphere of the sky. This small detachment of the stellar army is reinforced by the incalculably more numerous hordes that are revealed by the telescope and the further multitudes that evade the grasp of even the largest telescopes at present available.

Assuming that the night light was stellar in origin, the American astronomer Newcomb drew attention, more than half a century ago, to the interest attaching to the accurate determination of its intensity: such a determination would, according to his view, give us some idea of the total number of the stars — visible or, individually, invisible — that populate space all around us.

In itself, such a principle is valid. But the more recent statistics and accurate measures of Yntéma, Fabry, Dufay and many others (mostly provided by photography) have shown that the night light cannot be wholly accounted for in this way: it is too bright (if one may use the word in such a connection) to be provided by the stars alone. Calculations based on their numbers and their progressive decrease of magnitude have in fact shown that the stars contribute no more than 15 to 20% of the faint light which dissipates the darkness of a moonless night.

To account for the deficit, it is necessary to assume that the background of the sky itself is faintly luminous. The sources of this luminosity are various; one of them, the Zodiacal Light, is of great interest in its own right, as we shall see.

The Zodiacal Light. — This celestial phenomenon is little known, or even completely unknown, to the general public, for it does not present itself flamboyantly to our notice and its observations requires attentive examination by an informed observer. First of all, therefore, we had better describe its appearance and the manner of its presentation in the sky.

The Zodiacal Light appears as a very diffuse, faint patch of luminosity, in the shape of an immense cone (its true shape is that of a spindle — two cones placed together base to base) rising more or less obliquely into the sky above the fading twilight; alternatively, it is to be seen in a similar attitude above the eastern horizon before sunrise.

From its varying appearance and position the Zodiacal Light would appear to be a vast, very flattened zone of luminosity, centred in the Sun. Its name derives from the fact that its axis coincides approximately with the ecliptic,* which, it will be remembered, lies centrally within the band of the Zodiac.

Owing to this orientation, the visibility of the Light varies considerably from one season to another, and depends also on the observer's geographical position. The reader is here referred back to the explanations given in

* The great circle marking the annual path of the centre of the Sun's disc; it intersects the celestial equator at the equinoctial points, and is parallel to it at the solstices. See p. 48.

Fig. 547 — THE ZODIACAL LIGHT AS SEEN IN OUR LATITUDES AT THE END OF TWILIGHT IN SPRING

earlier chapters of the varying appearance of the star sphere, the mechanism of the seasons, the apparent diurnal and annual motions of the Sun, etc. From these accounts it will be realised that the orientation of the ecliptic relative to the horizon changes considerably from season to season at dusk and dawn — the times when the Zodiacal Light is visible.

This is illustrated in Figs. 551a and 551b, which show the Light's varying extent and its nightly duration of visibility from two representative zones of Earth's surface.

Fig 551a refers to its visibility in temperate latitudes of the northern hemisphere — such as those of Europe, for example. It will be seen that optimum observational conditions are obtained during the evenings at the spring equinox: the axis of the cone is then more nearly vertical to the horizon than at any other time, and therefore it suffers minimum interference from the horizon mists and the last traces of twilight. The contrary is true at the

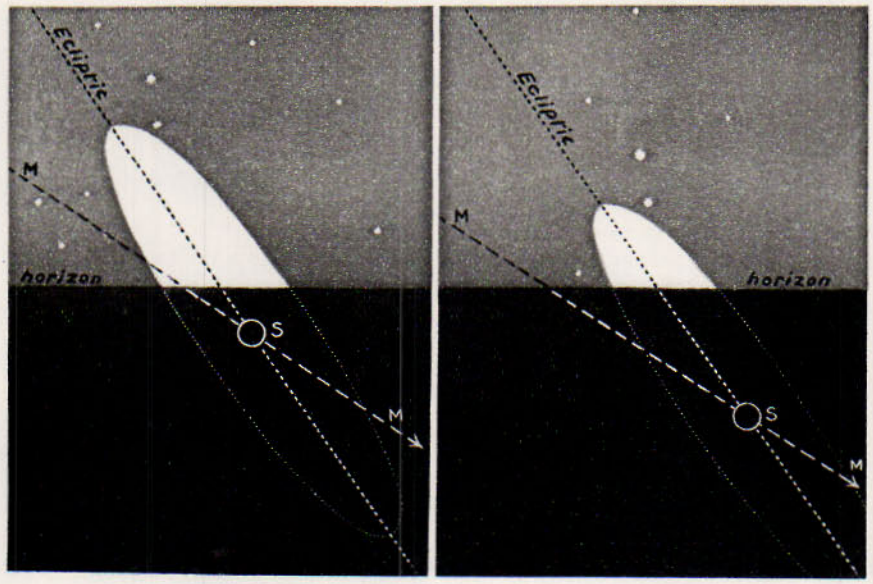


Fig. 548 — THE ZODIACAL LIGHT ASSUMES THE GENERAL SHAPE OF AN ELLIPSE CENTRED ON THE SUN, SINKING BELOW THE HORIZON AS NIGHT FALLS (SPRING)

Fig. 549 — THE ZODIACAL LIGHT PHOTOGRAPHED WITH A CONDENSING LENS FROM A MAGIC LANTERN

autumn equinox, when, owing to its inclination, the cone rapidly follows the Sun below the horizon. Intermediate conditions are encountered at the solstices, but the summer skies do not favour the observation of the Light owing to the prolongation of twilight, which retards its visibility. Thus in practice the Zodiacal Light is only visible in the evenings from temperate northern latitudes during the winter and the early spring. Conditions for its observation in the morning are the exact reverse: it is best seen from the autumn equinox on into the winter months.

Fig. 551b shows that in equatorial regions the evening cone (whence, also, the morning cone) is much more favourably presented at all seasons of the year. At the solstices it is almost vertical, and is only slightly inclined to the horizon at the equinoxes. In these regions, therefore, conditions for the observation of this delicate luminosity are as favourable as possible, and are still further improved by the short twilight and the greater atmospheric transparency of the tropics.

It is worthwhile emphasising these general conditions governing the visibility of the Zodiacal Light, so that readers who have never seen it may have the best chance of catching a glimpse of it. Its effective observation demands the utmost atmospheric clarity and (even more important) the absence of the Moon and any artificial lights whose glare may be diffused in the atmosphere or reduce, however slightly, the dark adaptation of the retina.

What we see is in fact only the brightest zone of the Zodiacal Light. The most cursory examination will show that it fades insensibly from its centre outwards, so that no definite 'edge' can be assigned to it; this is particularly true of the apex of the cone. It has furthermore been established that the Light fades more gradually into the background of the sky to the north than it does to the south. The brighter central region of the Light is from 25° to 30° wide at the horizon, while its height is at least three times as great. Sometimes, when the clarity of the atmosphere is exceptional, zones of decreasing intensity may be distinguished fairly clearly. Its form likewise seems to be variable, sometimes being abnormally wide or abnormally high, especially to the north. Lastly, its general brightness is subject to fluctuations.

What we are dealing with is in fact a sort of extension of the solar corona. The Zodiacal Light is now known to be sunlight scattered on interplanetary dust clustering around the ecliptic, and on free electrons in the same area; it is of course this last aspect of it which leads us to speak of it as an extension of the corona.

This is particularly relevant to the topic we have been discussing — the brightness of the night sky. According to the evidence, what we see is only the brighter, and denser region of the Light, which in reality extends all around the Earth in every direction; is should therefore impart a faint luminosity to the whole sky. Such a luminosity, as we have just seen, is in fact observed.

On nights of great atmospheric clarity — particularly in the tropics or from mountain summits — it can be seen that the main cone of the Light is prolonged in the form of an increasingly faint band as far as the antisolar point; here it brightens very slightly into a patch of light known as the *Gegenschein*, or Counterglow. The explanation of this appearance is to be found in the scattering, within a cloud of fine particles, of light coming from directly behind the observer.

The physical analysis of its radiation demonstrates clearly that the Zodiacal Light is composed of particles of this sort, but it does not necessarily follow that the above account of its form and distribution is correct. All the same, it is clear that its spatial distribution is in some way related to the Sun. In the first place, it is only approximately true to say that the Zodiacal Light, as seen from the Earth, coincides with the ecliptic. The consensus of very many observations is that it is in fact inclined at about 7° to the plane of the Earth's orbit, and is effectively coincident with the plane of the Sun's equator. Various manifestations of the Sun's activity, such as the greatest development of the corona in its equatorial regions, suggest that the production of the Zodiacal Light may be dependent on the Sun. Variations of its observed brightness, which can without difficulty be correlated with the solar cycle, lend plausibility to this view. On several occasions, also, an exceptional brightening of the Light has been observed; on one such occasion it was associated with an abnormal northward extension, to the accompaniment of an auroral display.

All this suggests that the Zodiacal Light is not so much a cloud of fine cosmic dust — the remains, according to

Fig. 550 — ASSOCIATED LIGHTS SOMETIMES APPEAR AT THE EDGE OF THE ZODIACAL LIGHT

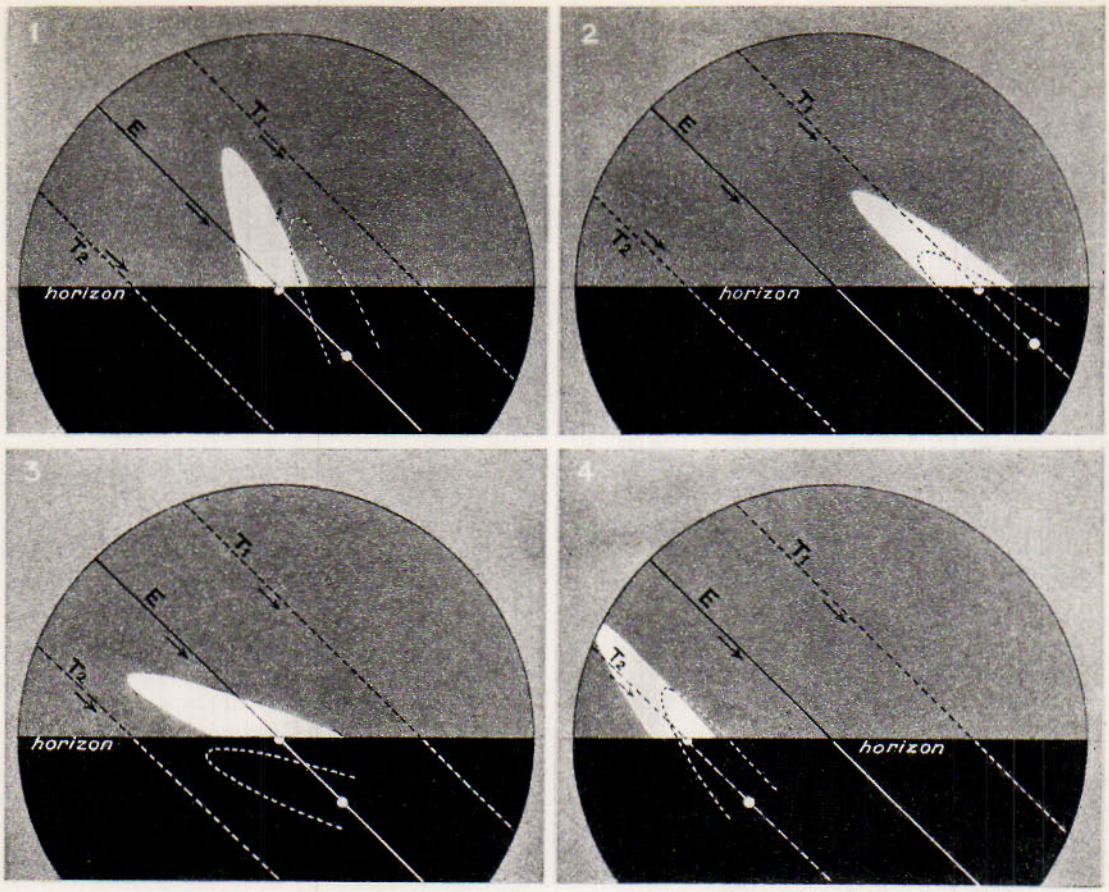


Fig. 551a — THE ZODIACAL LIGHT. Observing conditions in temperate latitudes during the evenings: 1, spring; 2, summer; 3, autumn; 4, winter.

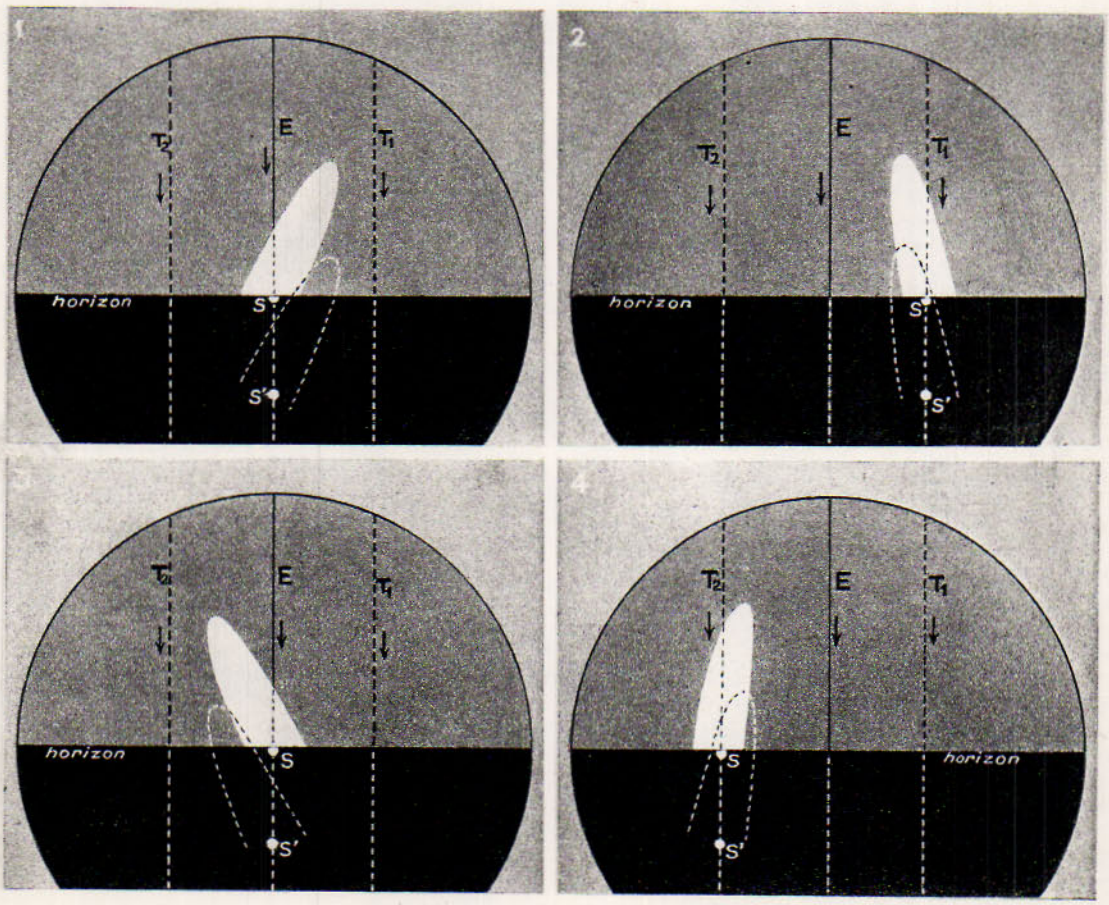


Fig. 551b — THE ZODIACAL LIGHT. Observing conditions at the equator during the evenings: 1, spring; 2, summer; 3, autumn; 4, winter.

Fig. 552 — NORTHERN HORIZON LIGHT (VISIBLE DURING THE SUMMER) which seems to be an extension (very considerable) of the Zodiacal Light.

some views, of a primeval nebula — as a product of the Sun's corpuscular radiation. In this case it would have to be regarded as a rarefied gas, composed for the most part of free electrons. At all events its constituent particles, whatever they may be, simply scatter the Sun's light without adding any of their own, since no emission lines have been detected in its spectrum.

One further question remains. Does this cloud occupy the whole of the void between the Earth and Sun, or is it concentrated in the vicinity of the Earth? The latter possibility is suggested by the various theories which envisage the capture of solar electrons by the Earth's magnetic field (see last chapter), and the consequent accumulation of these particles in the immediate neighbourhood of the Earth. This hypothesis has obvious points in common with that which accounts for the polar aurorae, thus adding weight to the idea that there may be some resemblance or connection between the two phenomena.

The Zodiacal Light still offers a fruitful field for observation and physical investigation. But whatever may be its true cause, this faint and enigmatic light contributes to the brightness of the night sky, for it extends more or less obviously over the whole of the star sphere. Its contribution has not yet been accurately evaluated, and in any case depends on the observer's geographical position, but we are entitled to put it at about 10 to 15% of the total recorded night light.

Permanent general aurora. — A part of the illumination of the night sky is produced, as we have seen above, in the Earth's own neighbourhood — more precisely, in its upper atmosphere — but though the source of this radiation may be near at hand, its ultimate cause must be sought far from the Earth's vicinity.

This phenomenon may be likened to a faint aurora which is not confined to the polar regions, but extends at high level throughout the whole of the Earth's atmosphere. Its existence has been revealed by the spectroscope — an achievement demanding the greatest technical finesse, owing to the faintness of the light and to the presence of other radiations, different in origin, from which it has to be distinguished. In general terms, this spectrum shows that the source of the radiation is twofold: one is an emission of light, and the other is merely a scattering of incident sunlight.

It has been demonstrated that certain of the elements comprising the upper layers of the Earth's atmosphere are emitting radiation. The observed spectrum is complex and its origin certainly multiple, for it contains certain lines which are characteristic of polar aurorae — the most intense being the green line of oxygen — but also others which serve to differentiate clearly between the two classes of phenomenon. Even though the production of the polar aurorae may be explained in the way already described, that of the general aurora, though similar to it, still remains undetermined. Several hypotheses have been proposed, all of which envisage the Sun as the ultimate cause. Thus Sommer attributes the general aurora to the dissociation of atmospheric zone; according to Dauvillier, it results from the excitation of atmospheric oxygen by solar cathode rays.

The work of Cabannes, Dufay and their colleagues — from which the principal results summarised here are derived — has thrown much light on the different altitudes at which the phenomenon occurs. Measures based on the increasing intensity of the light towards the horizon (i.e. with increasing thickness of the observed layer) indicate altitudes for the most part considerably greater than 60 miles. Recent results published by Dufay suggest that there are several luminous layers, between 40 and 60 miles, and again at about 140 miles. M. and Mme. Vassy conclude from their investigations that there are two thin layers, one at about 40 miles and the other and more intense at about 600 miles. The principal radiations have furthermore been identified as those of ordinary constituents of the atmosphere — nitrogen and oxygen, neutral or ionised, and (above about 60 miles) as frequently atomic as molecular.

The precise mechanism whereby these radiations are produced is still imperfectly understood, but it is sufficiently evident that this high-level luminescence contributes a great part of the brightness of the night sky — as much as 70%, according to the most recent estimates.*

Though the general aurora is a permanent feature of the upper atmosphere, its intensity nevertheless varies. In the first place, the atmospheric levels that are involved in its production are not homogeneous — according to Dufay, they should be thought of as an assemblage of luminous clouds, whose brightness changes during the course of the night. Obviously, too, allowance must be made for local meteorological conditions, which are capable of introducing serious discrepancies into the observations. But even when these have been taken into account, real and periodic variations of brightness remain; their cause would appear to lie in the physical properties of the atmospheric layer and in the intensity of the external excitation. Among these large-scale variations, some are annual and some semi-annual. Finally, it seems likely that an even longer variation is superimposed upon them, and that this can be correlated with the solar cycle. Furthermore, a factor which is certainly relevant is the position of the Earth in its orbit, since (the latter being slightly inclined to the plane of the Sun's equator) this will determine the degree to which the Earth is liable to bombardment by the electrified particles issuing, for the most part, from the most active zones of the Sun. Much interesting work is being done on the elucidation of this connection.

Thus it can be seen that this component of the night

* In October 1955 air glow was produced artificially, following a suggestion made by Bates in 1950. 3 kg. of metallic sodium were released from Aerobee rockets at heights of from 70 to 93 miles, and a marked increase in the intensity of the air glow emission in the yellow was recorded.

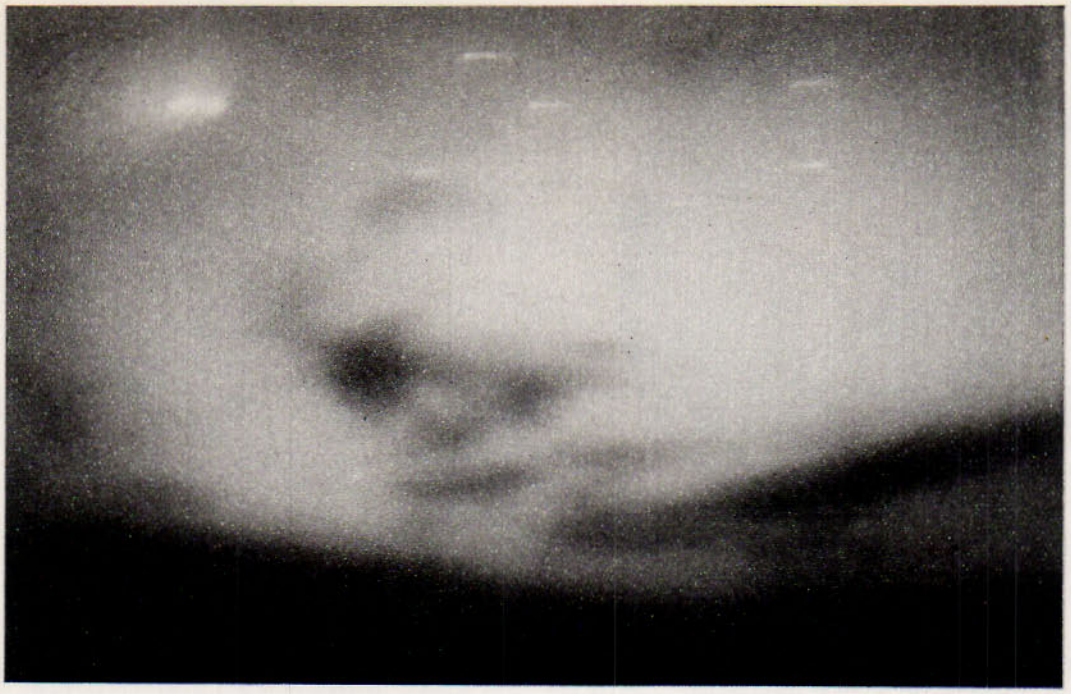


Fig. 553 — THE INTENSIFICATION OF THE LIGHT OF THE NIGHT SKY TOWARDS THE HORIZON ALLOWS THE SILHOUETTES OF CLOUDS TO BE DISTINGUISHED

light, although it is produced only a few miles away, within the Earth's own atmosphere, presents the astrophysicist with some nice problems. To answer other and still more pressing problems, we must now turn our attention to the immensities of interplanetary and interstellar space.

The scattering of light in space. — We have seen that the night light consists partly of radiations originating in the upper atmosphere, and partly of reflected (or, more accurately, scattered) sunlight and, possibly, general starlight. Since sunlight cannot directly reach the night-side of the Earth it must be scattered by something far outside the Earth's atmosphere, in the depths of space.

We have seen, too, that the Earth is immersed in the Zodiacal Light, but that this can account for only a small fraction of the total observed luminiscence. To locate the origin of the remainder we must look even further afield.

Space is not the perfect vacuum it was once thought to be: it has, in fact, a very numerous and varied content. This consists, first, of solid bodies, for the most part microscopic, moving in every conceivable direction, which become 'shooting stars' or meteors when they enter the Earth's atmosphere. Apart from this single circumstance they emit no light of their own, but nevertheless reflect sunlight just as the planets do. Their number, though unknown, is certainly very large. Secondly, there is the infinitely greater number of individual molecules, atoms and free electrons, which have escaped from the atmospheres of the planets, from comets, and above all from the Sun, and which constitute a sort of ultra-rarefied gas. It is true that the medium formed by all these various particles is of the utmost tenuity, but no matter how weak their scattering of the Sun's light may be, by virtue of the fantastic volume of space that they occupy, with consequent piling-up in depth along the line of sight, they ultimately succeed in producing an appreciable fog of light.

A great deal still remains to be discovered about the interplanetary cloud, both qualitatively and quantitatively, but at least it is recognised that it contributes effectively to the overall luminosity of the night sky.

Even with the vast distances of interplanetary space, however, we have not reached the frontiers of our enquiry.

Interstellar space. — Interplanetary space, and especially the vicinity of the Sun, thus seems to be stocked with a rich variety of miscellaneous matter. But this state of affairs is not limited to the Solar System: to some extent* — a very variable extent — a similar medium is to be encountered throughout the enormous expanses of space occupied by the Sun and its stellar companions.

The stars are in fact immersed in what is known as the cosmic cloud, like lights in a very thin fog. We shall be more concerned with this impalpable 'fog' — whose very existence remained unsuspected until fairly recent times — when we come to the chapter dealing with the star system as a whole, the Galaxy. But we may say in anticipation of this that we are able to demonstrate its existence by direct observation — at any rate in some regions. It is indicated by the existence of nebulae, objects of the most diverse and fantastic shapes, which are now recognised to be no more than local condensations of the cosmic cloud.

This corpuscular and gaseous population of space also co-operates in dissipating the darkness of night, for the light of the stars scattered by the interstellar medium is by no means negligible. In some regions, however, it plays exactly the opposite rôle: that of an obscuring screen or filter. We shall learn later how the interposition of this kind of fog, if in sufficient depth, can partially or even completely absorb the light of more distant stars. Because of interstellar absorption, the earlier estimates of the distances of such stars have had to be revised. In other words, they are less distant than they would be if faintness were solely a function of distance.

Once we leave the Solar System we are involved with distances so enormous that it would be well to familiarise ourselves with them straight away, so that we shall be

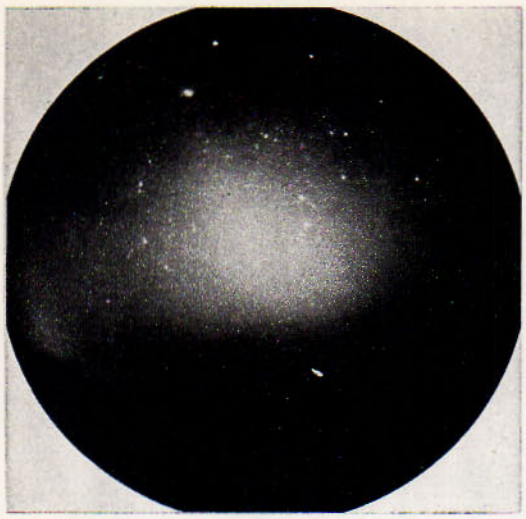


Fig. 554 — THE ZODIACAL LIGHT ON A SPRING EVENING. Near the top, the Pleiades can be seen; close to the horizon, below a belt of cloud, Venus.

able to form some concrete idea of the immensity of the stellar universe. The numerical expression of stellar distances tends to be more astonishing than informative, for the very reason that they so vastly transcend the scale of our everyday experience. We have already seen that an attempt has been made to give more meaning and significance to these distances by employing a new unit of measurement: the time that light takes to cover them. And, indeed, the duration of the journey of a ray of light travelling at 186,000 m.p.s. does convey in a striking and meaningful manner an idea of the length of that journey.

The dimensions and distances that we have already encountered within the Solar System may seem to be enormous enough. Apart from the Moon and the other satellites, which are relatively close to the planets about which they revolve, the miles pile up in their millions, hundreds of millions, and even thousands of millions as we approach the outskirts of the System.

It is mere child's play for a ray of light to cross the Solar System: only 8m 18s after it has left the Sun it has reached the Earth, and within 7 hours it will have crossed the orbit of Pluto and left the System altogether. But to reach the nearest star, it will have to continue its journey for over 4 years. This gives an impressive idea of the vastness of the interstellar void, for the distances separating neighbouring stars are everywhere of this general order.

Thus the size of the Solar System is negligible compared with the wastes of space that surround it. And if, as we have reason to believe, other stars are accompanied by planetary systems, more or less similar to our own, then these may be picturesquely likened to tiny villages, each at the centre of its own Sahara desert.

Even more astounding, perhaps, is the comparison between the sizes of the stars and the distances separating them. The average star is no larger than a single buckshot, a millimetre in diameter, situated at the centre of a cube of empty space, the length of whose sides is 125 miles.

It should therefore be realised that however large the stars may appear to be by our terrestrial standards — and the majority exceed a million miles in diameter — they are none the less negligibly microscopic compared with the void that surrounds them on all sides. This void, in truth, we can no longer think of as a perfect vacuum. The interstellar medium, despite its tenuity and because of its enormous extent, represents a very considerable quantity of matter.

We have now reached the point where we must leave the Sun's domain and launch out among the stars.

Fig. 555 — THE ZODIACAL LIGHT SOMETIMES EXHIBITS FAIRLY WELL DEFINED ZONES OF VARYING BRIGHTNESS

BOOK III

THE REALM OF THE STARS

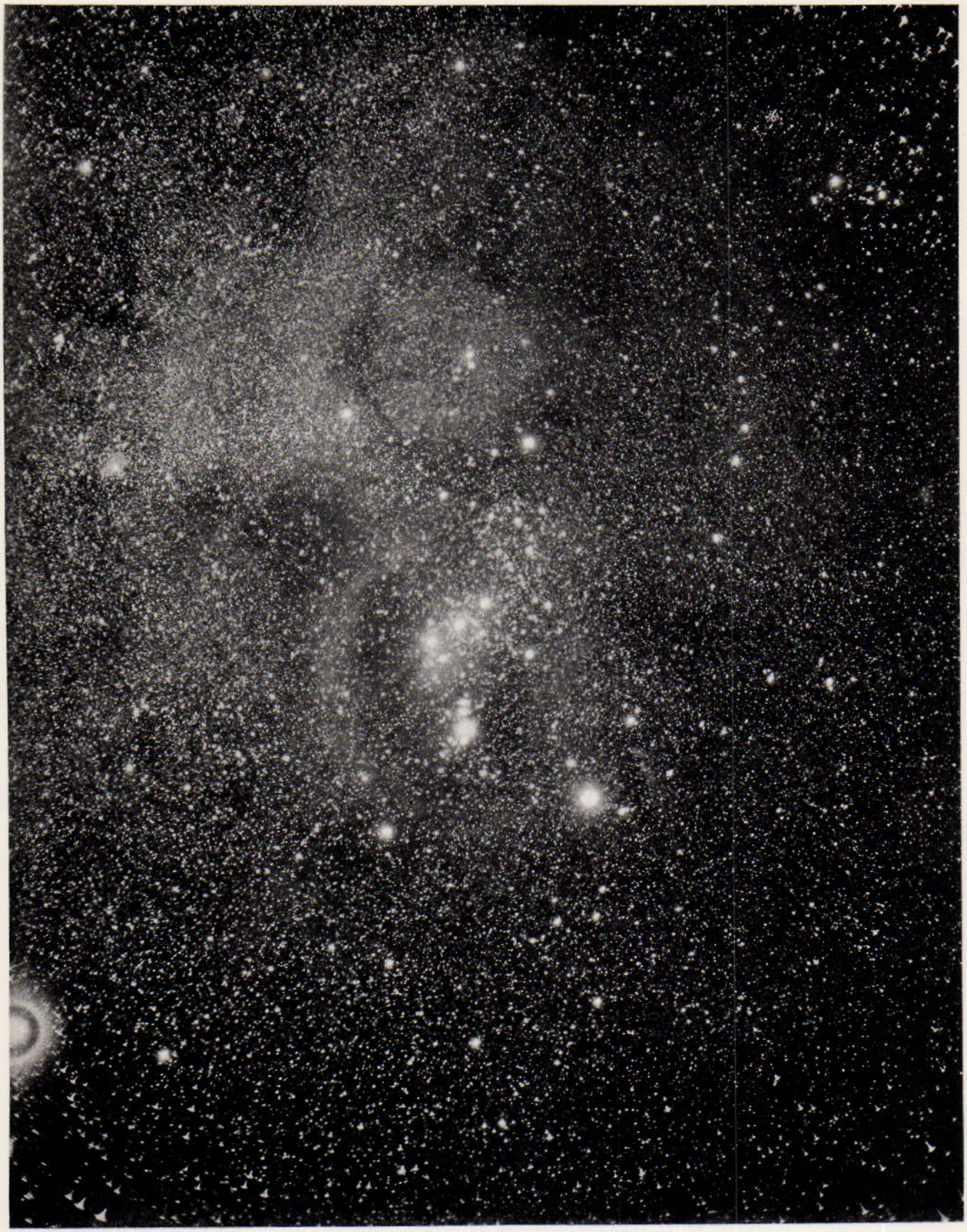


Fig. 556 — ORION, THE MOST BEAUTIFUL CONSTELLATION IN THE SKY, WITH SIRIUS (LOWER LEFT), THE BRIGHTEST OF THE STARS. 10-hour exposure with a wide-angle Tessar lens. *(Photo, Mount Wilson Observatory)*

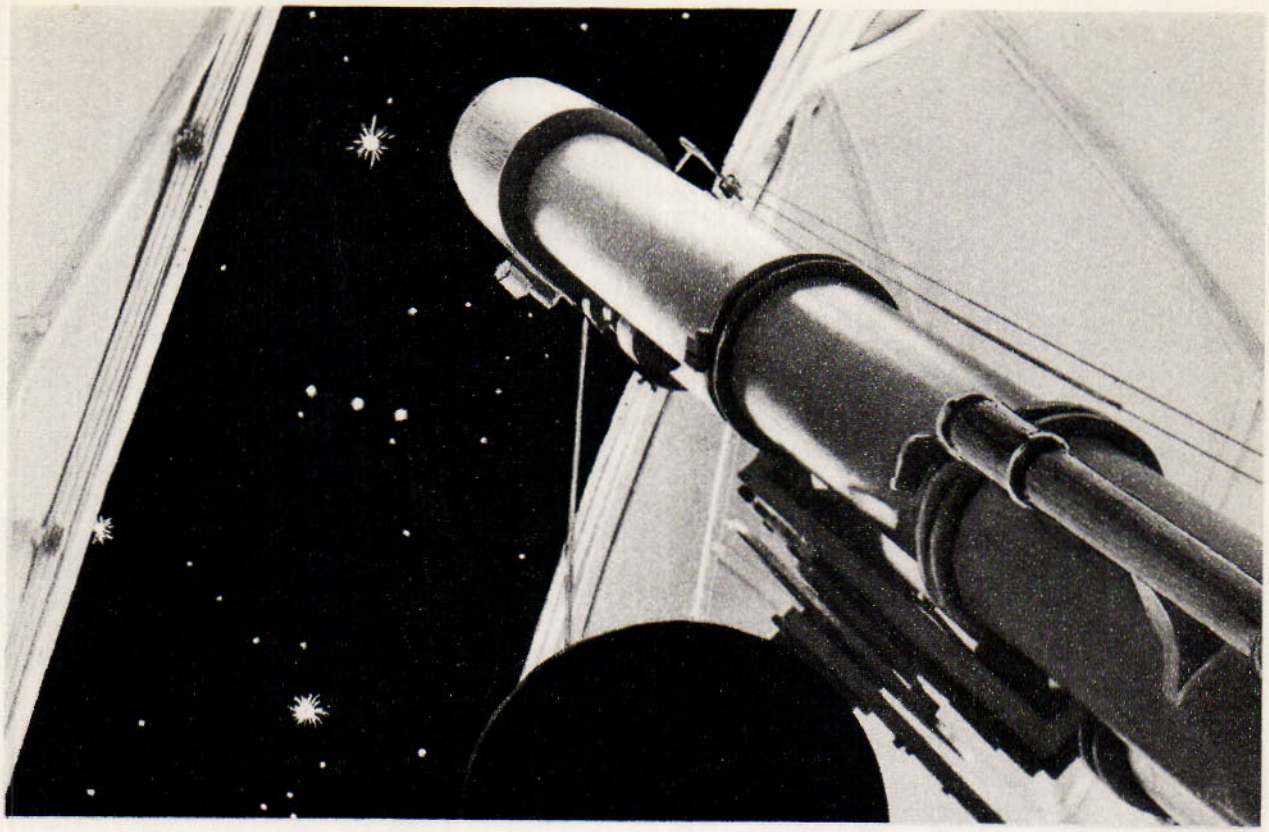


Fig. 557 — THE CONSTELLATION OF ORION SEEN WITH THE NAKED EYE

CHAPTER 12

THE STARS IN THE SKY

We must now leave the Solar System and turn our attention to the stars, whose teeming myriads populate space in every direction. As a spectacle, the star-glittering night sky is undoubtedly beautiful, but until quite recent times the stars appeared to astronomers to be less fruitful as a field for observation than our more immediate celestial neighbours. A star is a mere point of light; and ordinary vision, even when reinforced by the most powerful telescopes, can tell us nothing about its nature. Furthermore, observation of the stars was for a very long time limited to general studies, principally of a statistical nature, which in certain respects are less interesting than the visual study of the more humble members of our own Solar System, which readily lend themselves to descriptive observation of the most intriguing kind.

But thanks to the ingenious observational techniques of modern astronomy we can now investigate the real dimension of the stars and their individual chemical and physical characteristics, as well as their previously unimagined — even unimaginable — numbers, and their distribution throughout a Universe whose immensity numbs the human mind.

Stellar astronomy has developed so rapidly that in many respects it now dominates all other branches of the science. But the absorbing interest of the more recent investigations should not make us forget the starting point of these enquiries: the solitary contemplation of the starry sky which everyone must make for himself.

The constellations; star charts and catalogues. — When the star sphere is examined on a really clear night, two things chiefly stand out: the vast numbers of the stars of every grade of brightness, and their arrangement in groups whose patterns are striking enough to catch the eye — the so-called constellations.

For long ages mankind had to be content with these two elementary facts, of which the first falls far short of the truth and the latter is completely illusory: for it is only an effect of perspective that makes stars which are situated at very unequal distances appear to be near to one another when projected on the same plane, that of the imaginary star sphere.

Nevertheless, these elementary observations (which are intimately connected with one another, since the patterns of the constellations are determined by the distribution of the visible stars) are the foundations on which our knowledge and catalogues of the heavens are based. The constellations also serve a useful practical purpose: they form a pattern by reference to which one can orientate oneself and specify immediately the position of anything one wishes to indicate. We have already seen how astronomers have covered the whole star sphere with

a system of co-ordinates which define any given position in terms of Right Ascension and Declination (equivalent, respectively, to terrestrial longitude and latitude). A network of imaginary lines is thus superimposed on each constellation, and the region of the sky that it occupies can be defined by the co-ordinates that bound it.

We know also that the stars provide a virtually fixed background against which are projected such nearer bodies as the Moon, planets and comets; these, by virtue of their motions, appear to move among the stars. Their successive future positions on the star sphere can be calculated and expressed in terms of their co-ordinates, R.A. and Dec. A statement of the constellation they are in is, however, more immediately informative; thus if we want to observe Saturn and find from the ephemeris that its position is R.A. 7h, Dec. $+23°$ we shall then have to refer to a star atlas to discover where this position actually is on the star sphere (unless we are extremely conversant with the use of astronomical co-ordinates); on the other hand, if it is also stated that Saturn is in Gemini, even an observer who is not very familiar with the general appearance of the star sphere will probably know at once where to look. The same applies to the positions of novae, whilst the paths of meteors are also defined with reference to the fixed stars. Quite apart from their aesthetic attributes, therefore, the constellations are helpful from many points of view, and for this reason their use has been retained by modern astronomers.

Some constellations are particularly noticeable, either because of their striking pattern or because of the brightness of their stars; others are less eye-catching. In all, there are 89 constellations, and between them they cover the whole star sphere. Their distribution is as follows:

(a) Round the ecliptic are the 12 constellations which have given their names to the Signs of the Zodiac; but owing to the precession of the equinoxes a constellation and a Sign of the same name no longer coincide. A part of the constellation Ophiuchus intrudes between two of them, Scorpio and Sagittarius, so that in reality there are 13 ecliptic constellations.

(b) 29 constellations between the ecliptic zone and the north celestial pole, plus the remainder of Ophiuchus.

(c) 47 constellations between the ecliptic zone and the south celestial pole.

When one looks at the night sky one's attention is first attracted by those star groups which are composed of comparatively bright stars. But the regions between these more noticeable groups are also populated with stars, though they may be fainter. The boundaries of each constellation therefore enclose a very much more extensive area of the star sphere than that occupied by the configuration of its brighter stars. Most of these boundaries were assigned quite arbitrarily by the ancients.

The names of most of the constellations date back to the times of antiquity, and refer to mythological or legendary men, women, and animals. These identifications had an unfortunate influence on the design of celestial globes and star charts of earlier times: the positions of the stars might indeed be shown with the greatest accuracy then possible, but they would be covered by drawings (often in exaggerated detail) of the human figures, animals, etc., which give their names to the various constellations. Moreover, it would seem that in many cases the outlines of the constellations were distorted to make a closer fit with the more or less complicated figures which they were supposed to represent. This method of depicting the constellations may perhaps have had a mnemonic value, but it certainly reduced the documentary value of the maps by obscuring the real patterns of the star groups.

Less artistically, modern star charts confine themselves to connecting the principal stars with straight lines so as to emphasise the salient shape of each constellation. Finally, the capricious constellation boundaries of the ancients have been replaced by a geometrical system of delimiting the constellations, all of whose boundaries are now either meridians of Right Ascension or parallels of Declination.

Very early in the history of astronomy it became necessary to give names not only to the constellations but to individual stars. At first, the brightest stars were given proper names, many of which, Arabic in origin, survive to the present day — Sirius, Altair, Vega, Aldebaran, Rigel, etc. Alternatively, each star was designated by a description of its position in the figure that the constellation was imagined to represent. This system of nomenclature, whose complexity and prolixity can be imagined, was replaced by the simple and rational method of designation by means of the letters of the Greek alphabet which was introduced by Bayer in 1613. In each constellation, the brightest star is α (alpha), the second brightest β (beta), and so on through the alphabet. But owing to the richness of the starry background in some constellations, and the size of others, this series of letters was quickly exhausted; it was then followed by the letters of the Latin alphabet, and finally by numbers (see Fig. 558).

This nomenclature was in theory only used for the naked-eye stars, however, and the requirements of a growing science conflicted with an ever more crowded inventory.

General star charts need to be supplemented by more detailed information, such as the accurate determinations of the positions and brightness of the stars, so that the compilation of catalogues of stars, containing all these data, was a natural development.

As a matter of historical interest, very simple lists of this kind are known to date back to the times of the Babylonians, but the first true catalogue, so far as we now know, was that of Hipparchus (127 B.C.) This, unfortunately, has not survived, and the earliest extant catalogue is the *Almagest* of Ptolemy (A.D. 137), containing 1028 stars. To this Alexandrian astronomer we also owe the classification of the stars by magnitudes, according to their apparent brightness (see p. 304).

Later catalogues, compiled by the Arab astronomers during the Middle Ages and by Tycho Brahe in 1594, were no more comprehensive than the *Almagest*. Moreover they only included the brighter naked-eye stars, which at the time were all that were known. It was not until Galileo first applied the telescope to the study of the skies (1610) that the true richness of the starry background — which increases with every increase of telescopic power — began to be appreciated. By the end of the 18th century, William Herschel's great reflector was already enabling him to count many thousands of stars per square degree in certain regions.

The need for catalogues of the greatest possible completeness and accuracy became more and more urgent as astronomy advanced. The improvements both of instruments and of methods of measurement that were made during the 17th, and still more during the 18th century, enabled these to be undertaken. Among the more important of these impressive monuments to human patience, which mark the opening of the modern era of astronomy, the following should certainly be mentioned: Lacaille's catalogue of 10,000 southern stars (1757); Bradley's (Greenwich, 1763), long after referred to by workers in the field of stellar proper motions; Lalande's (Paris, 1801), giving the positions of about 47,000 stars; and, during the 19th century, Argenlander's great catalogue of 324,000

stars (published at Bonn between 1852 and 1862) known as the *Bonner Durchmusterung* (abbreviated, *B.D.*). This was completed by the addition of two supplements: a catalogue of 134,000 stars south of Declination −23° which was published in 1886, and the *Cordoba Durchmusterung* (*C.D.*), giving the positions of 570,000 still more southerly stars.

Nowadays, then, stars are designated by Greek or Roman letters if they are bright enough to have figured in the old maps (these include most of the stars visible to the naked eye), or by their number in such or such a catalogue preceded either by the catalogue's recognised abbreviation or by the name of its compiler. We shall encounter numerous examples as we go along.

Still other catalogues are devoted to some particular type of star — such as double stars or variable stars (see Chapter 13). But those already mentioned will suffice to indicate the magnitude of a task which consists in building up an inventory of the whole sky and determining the positions, one by one, of hundreds of thousands of stars, each requiring the most meticulous measurement.

Yet the stars that figure in all the catalogues that have ever been made represent only an infinitesimal fraction of the total number. To tackle these by ordinary visual methods of observation would be an impossible task, and only photography, which reveals countless myriads of stars beyond the reach of the visual telescope, can hope to do any better. The photographs reproduced here are more eloquent of this than any words.

Among the earliest attempts to apply photography to the task of recording the stars was that of the brothers Henry at the Paris Observatory, around 1885. The remarkable photographs they obtained of various regions of the heavens inspired Admiral Mouchez, at that time Director of France's National Observatory, to propose

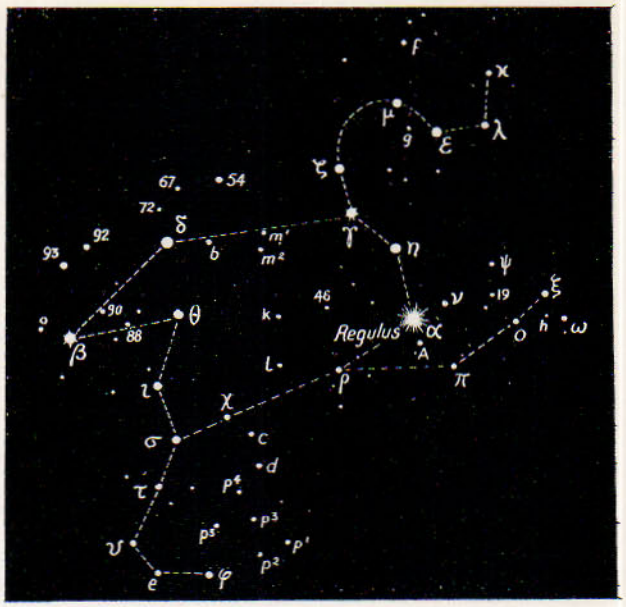


Fig. 558 — A CONSTELLATION: LEO

the creation of a general photographic chart of the heavens. This ambitious project of course demanded international co-operation if the whole of the star sphere were to be covered. At the invitation of France, a conference was held in 1889 to study the proposal. Eighteen observatories, in all parts of the world, decided to co-operate, and to each was assigned a particular area of the sky.

This gigantic undertaking is still not completed,

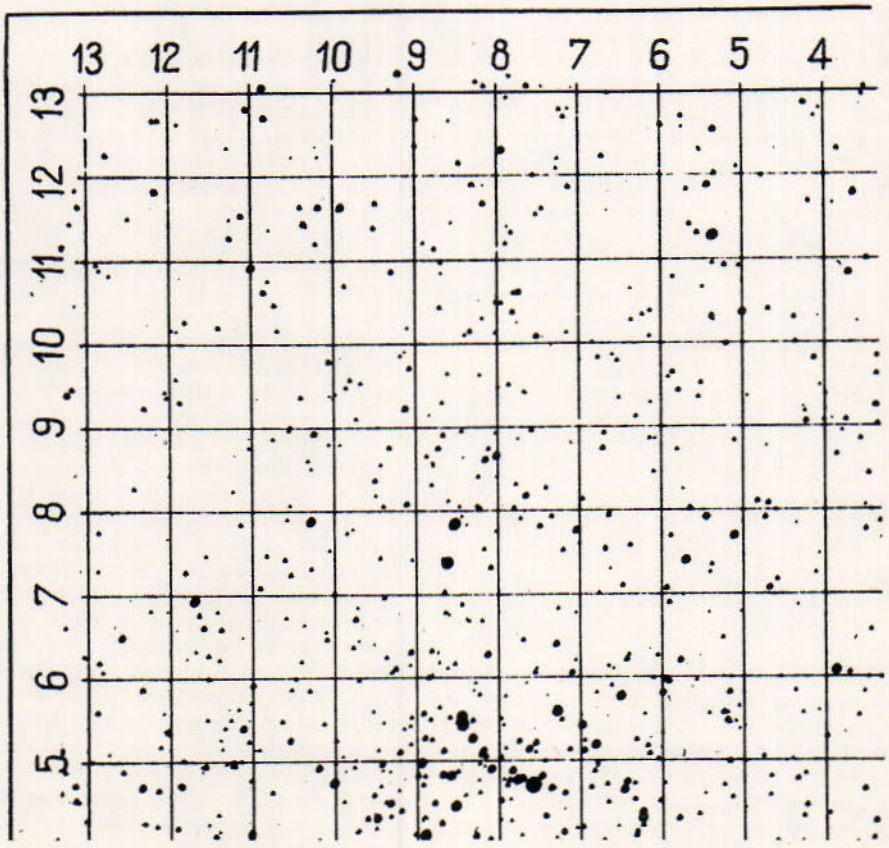


Fig. 559 — PART OF ONE PLATE OF THE CARTE DU CIEL. Each star is represented by three images, very close to one another, each of which is exposed for 30 minutes. *(Paris Observatory)*

though the publication of the *Carte photographique du Ciel* and its accompanying *Astrographic Catalogue* is now well advanced. It is estimated that the positions of ten million stars, measured from the photographic plates, have thus been accurately determined.

Since the initiation of the *Carte du Ciel*, instrumental advances have permitted the sounding of ever greater depths of space, with the discovery of so vast a number of faint stars that their comprehensive study and cataloguing are far beyond the bounds of the practicable. They can, however, be treated statistically, and at the suggestion of the Dutch astronomer Kapteyn, in 1905, 206 small areas of the sky were chosen; they were evenly distributed over the whole star sphere and each measured about one square degree. Each of these Selected Areas was then exhaustively examined and a catalogue compiled of its stellar population down to the limits that modern instruments can reach. The resulting combined catalogue of the Selected Areas has proved to be of inestimable value.

Lastly, there are the photographic atlases — no less useful, and indeed particularly informative if one is concerned only with the general distribution of the fantastic hordes of the stellar universe.

Apparent brightness and colour of the stars. — The unequal brightness of the stars is obvious at first glance: from the most brilliant down to the faintest that can be seen with the naked eye, every grade of brightness is represented.

Ptolemy was the originator of the classification of all the stars bright enough to be seen with the naked eye into six groups, or magnitudes, the brightest belonging to the first magnitude and the faintest to the sixth. This system has been retained to the present day, though the comparatively approximate eye-estimation of the various magnitudes has been replaced by methods which provide accurate photometric measures. Each magnitude is 2.5 times as bright as the one below it: thus a second magnitude star is 2.5 times fainter than one of the first magnitude and 2.5 times brighter than one of the third, and so on down the scale. If we represent the brightness of a first magnitude star by the value 100, then the relative brightnesses of the first 6 magnitudes will be as 100 : 40 : 16 : 6.3 : 2.5 : 1. In other words, the faintest stars that can be seen with the naked eye are 100 times fainter than those of the first magnitude.

The series is continued for the increasingly faint stars that are visible only with a telescope — magnitude 7, magnitude 8, etc. The following table quotes the faintest magnitudes normally visible (at mean altitude, and given clear skies), as a function of the diameter of the telescopic objective employed:

DIAMETER IN INCHES	MINIMUM VISUAL MAGNITUDE	DIAMETER IN INCHES	MINIMUM VISUAL MAGNITUDE
1.2	10	12	15
2.0	11	20	16
4.0	12	40	17
6.0	13	60	18
8.0	14	98	19

The present limit for visual observation is about magnitude 19, but owing to the cumulative nature of the formation of the photographic image, it can be extended to about magnitude 21 by photography. Stars of this magnitude are a *hundred million* times fainter than those of magnitude 1 — an impressive indication of the instrumental power that is available to the modern astronomer.

The outlines of this classification, once established, needed to be filled in, for the stars do not jump abruptly from one magnitude to the next. In other words between stars of magnitude 2 and magnitude 3, for instance, there are others exhibiting every intermediate degree of brightness. For these, decimal subdivisions of a magnitude are used, and they are graded as being of magnitude 2.1, 2.2, 2.3 . . . 2.9.

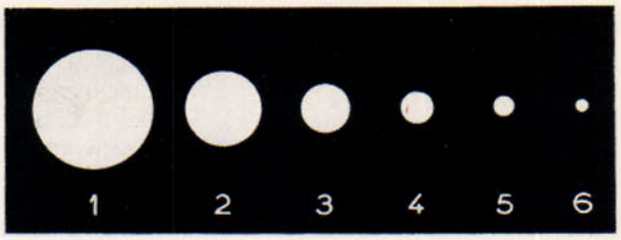


Fig. 560 — The areas of the circles are proportional to the brightnesses of stars of the first 6 magnitudes (i.e. those visible to the naked eye).

In the same way, the magnitude scale has had to be extended so as to include the brightest stars of all. These were originally all lumped together as 'magnitude 1', but in fact their brightnesses differ by a larger factor than 2.5 — i. e. their spread is greater than that covered by a single magnitude. Hence the original scale has had to be extended in the opposite direction, by the introduction of a zero magnitude followed by negative magnitudes — thus stars which are one magnitude brighter than those of magnitude 1 are counted as magnitude 0; those two magnitudes brighter than magnitude 1, as magnitude −1; three magnitudes brighter, as −2; and so on. Each of these negative magnitudes is decimally subdivided in the same way as before. The category originally called 'first magnitude' includes Capella (α Aurigae), which is intermediate between magnitudes 0 and 1 (actually, 0.2); Procyon (α Canis Minoris), slightly fainter, magnitude 0.5; Canopus (α Carinae), — 0.9; and Sirius (α Canis Majoris), −1.4. Sirius, the brightest star in the sky, is thus 1,000 times brighter than the 6th-magnitude stars which lie at the limit of naked-eye visibility.

In passing, it may be noticed that the same magnitude scale is used to express the apparent brightness of the planets, which appear to the naked eye as stellar points, though often surpassing these in brilliance. At maximum brightness Venus attains magnitude −4, Mars and Jupiter −3, Saturn −0.5; Uranus, on the other hand, only just visible to the naked eye, is never brighter than magnitude 5.7, while Neptune appears as a telescopic star of the eighth magnitude; as for Pluto, its magnitude of 15 limits its observation to telescopes of 12 in. aperture or more, although it becomes somewhat brighter near perihelion.

An important distinction must be made at this point. The magnitudes that we have been discussing refer solely to the apparent visual brightness of the stars. Most frequently, however, it is by means of photography that the brightness of the stars (and particularly of the fainter stars) is measured. Now it is well known to every photographer that the silver emulsion does not 'see' different colours in the same way as the human eye does: the retina is particularly sensitive to green, yellow and red, whereas the photographic emulsion is relatively insensitive to this colour range but is strongly affected by the blue, violet and ultra-violet radiation which the human eye only sees with comparative difficulty, if at all.

A glance at the night sky will show that the stars are not all the same colour: some have the bluish tint of an electric spark, many are white or yellowish like the Sun, others are of an obviously orange tint, while others again

are a definite red, glowing like rubies against the dark background of the sky.

It follows that the former, whose light is rich in blue and violet radiations, will affect the photographic plate strongly, while the latter, whose radiations lie primarily within the red and yellow range, will affect it only feebly. Thus if one attributes the same visual and photographic magnitudes to white stars, then bluish and reddish stars of the same visual magnitude will be respectively brighter and fainter on the photographic scale.

This difference between the visual and photographic magnitude of a star is known as its colour index, and we shall see later that it is a function of the star's temperature. Catalogues have been compiled of both visual nad photographic magnitudes.

A difference similar to, though less marked than, that between the reactions of the retina and the photographic plate to different colours, is to be encountered between the eyes of different observers, and even between the two eyes of the same observer, as well as between different types of photographic plate. This depends on small individual variations in the sensitivity of the retina or the emulsion to light of different wavelengths.

The resulting differences between the magnitude estimates of variously coloured stars made by different observers are known as their colour equation, or personal equation. Systematic discrepancies also occur between the magnitude measurements made at different observatories. For this reason a series of stars, whose brightnesses have been determined with the greatest possible precision so that their magnitudes are settled once for all, has been chosen to serve as a standard against which all other magnitude determinations may be compared. These stars lie in the vicinity of the north celestial pole and are known as the Polar Sequence; since 1922 this has served to define the universally adopted systems of visual and photographic magnitudes.

It should also be mentioned that by means of orthochromatic or panchromatic plates, used in conjunction with the appropriate yellow filter, it is possilbe to determine the visual magnitude of a star photographically; magnitudes determined in this way are known as photovisual magnitudes.

Lastly, a general point that should be emphasised: the stars cannot be observed save through the thickness of the terrestrial atmosphere, which absorbs some of their light; this effect, as has already been explained, is a function of the length of the light-path within the atmosphere. Hence the brightness of a star is considerably reduced when it is near the horizon, at rising or setting. At greater altitudes than about 45°, on the other hand, the absorption may be regarded as negligible. The values given in this book therefore refer to the stars' apparent brightnesses when they are observed at altitudes greater than 45°; their progressive diminution at lower altitudes is shown in the following table:

ALTITUDE ABOVE HORIZON	43°	32°	26°	21°	19°	17°	15°	13°	11°	10°	6°	4°
REDUCTION OF BRIGHTNESS (VIS. MAGS.)	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.5	2.0

Consider, for example, a magnitude 1 star such as Aldebaran: observed at an altitude of 19° it will appear as magnitude 1.5, and at 10° only 2.0. Sirius, the brightest star in the sky, can never be seen at its true brightness in our latitudes; owing to its position in the south celestial hemisphere it can never attain a greater altitude than 29° above the horizon of London. Hence this −1.4 magnitude star appears to be no brighter than −1.1 or −1.2, even when it is on the meridian.

These data can nevertheless only be regarded as rough indications, since atmospheric absorption affects different wavelengths unequally; the figures quoted above are approximately doubled in the case of photographic absorption, besides varying slightly according to the colour of a star. Thus the diminution is less marked in the case of a red star than of a blue one at the same altitude, owing to the selective absorption of the atmosphere which affects the short wavelengths (blue, violet and ultra-violet) more than the long. Every magnitude determination has also to be corrected for this selective effect.

These remarks refer, furthermore, to the basic absorption which occurs in an atmosphere of perfect clarity, In the vicinity of the horizon an additional factor, variable and unpredictable, comes into play – the haze and impurities which render the lower atmospheric levels relatively opaque, producing what is in effect a zone of extinction where the fainter stars are concerned. It is in fact generally recognised that one cannot see the stars until they have reached a certain minimum height (depending on the state of the sky) above the horizon.

From all this it will be realised that the accurate determination of stellar magnitudes is a process of very considerable delicacy and it will not be a matter for surprise that one occasionally comes across noticeable discrepancies between the figures given by different authors or quoted in different catalogues.

The numbers of the stars. – At first glance the sky, observed with the naked eye, would appear to be dotted with a myriad of twinkling points of light. This apparent richness of the stellar population is an illusion, probably produced by the splendour of the spectacle that is spread out above us. In fact, no more than 3,200 stars are to be seen with the naked eye on a perfectly clear night; and since only half the star sphere is ever visible at one time, this means that the total number of stars visible to the naked eye stands at something like 6,500.

We have just seen how the stars are graded according to brightness, the most brilliant belonging to magnitude 1; 20 stars can be counted which are brighter than magnitude 1.5: these are usually referred to, not quite accurately, as 'first magnitude' stars. But the stellar population increases rapidly with decreasing brightness, and there are, in round numbers, 50 stars of the second magnitude, 150 of the third, 500 of the fourth, 1,500 of the fifth, and rather more than 4,000 of the sixth. The progression continues with the fainter stars that are revealed by increasing telescopic power.

This increasing richness is easily explained. If we increase our visual penetration of space 10 times, for instance, the volume that can be explored is increased 1,000 times; in other word, if we assume that the stars are evenly distributed through space, their numbers will increase with the cube of their distance. This is only true, however, of the brighter and relatively near stars; for the fainter – and therefore, on the average, more distant – stars, the progression is more rapid still. We shall see later that star counts have provided valuable information regarding the spatial distribution of the stars. Meanwhile, we will content ourselves with quoting the figures of the American astronomer Seares, showing the total number of stars in the sky above various magnitude levels. (We have already seen how the definition of 'magnitude' is unfortunate from the point of view of linguistic convenience because magnitudes are represented by increasingly large numbers as the brightnesses become smaller.)

PHOTOGRAPHIC MAGNITUDE LIMIT	TOTAL NUMBER OF STARS
6	3,000
8	23,000
10	165,000
12	1,100,000
14	6,500,000
16	20,900,000
18	142,000,000
20	506,000,000

The total number of stars photographically accessible with existing instruments (i.e. down to magnitude 21) is in the neighbourhood of a thousand million. Even this, as we shall see in due course, is no more than a tiny fraction of the total number which make up our stellar system.

Stellar motions. — Our knowledge of the motions of the stars is comparatively recent. They are not apparent to the naked eye, and for this reason escaped the notice of the astronomers of antiquity, from whom the name 'fixed stars' derives. And it is quite true that the stars do appear to be immutably fixed to the star sphere: we see today the same constellations that circled above the heads of the Babylonian astronomers.

Yet the stars are everywhere hurtling through space like fantastic projectiles with velocities which, though they vary widely, are always very large by terrestrial standards. The reason for their apparent immobility is, once again, the scale of the distances that separate them from us, which has the effect of reducing their enormous linear displacements to nothing. Hence the sweep of their motions escapes immediate notice, and it is only after long lapses of time that their angular displacements on the star sphere become large enough to be perceptible.

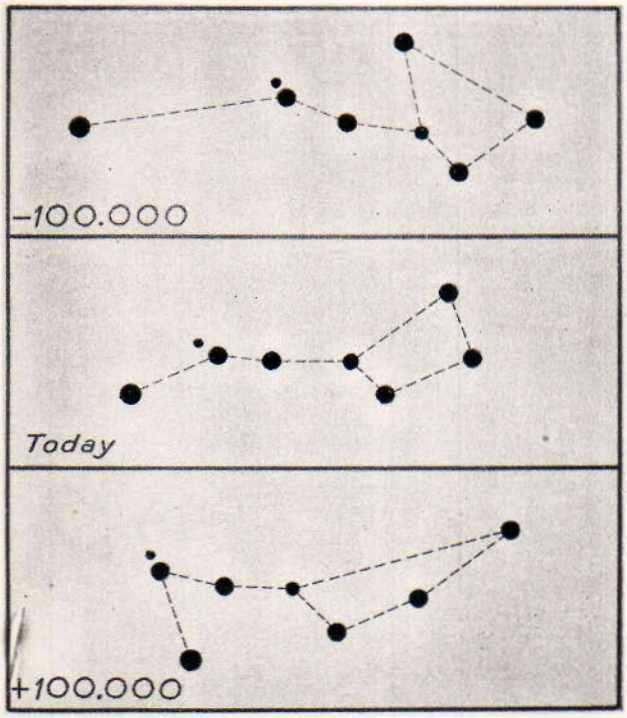


Fig. 561 — THE SLOWLY CHANGING SHAPE OF THE GREAT BEAR, DUE TO THE PROPER MOTIONS OF ITS STARS

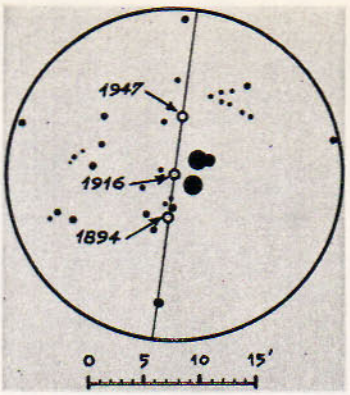


Fig. 562 — CHART SHOWING THE RAPID DISPLACEMENT OF BARNARD'S STAR BETWEEN 1894 AND 1947

The measurement of these demands the utmost precision and powerful instrumental assistance. Even so, it is only the nearer stars whose displacements can be detected at all — and then only when their linear velocities are high, and directed more or less at right angles to the line of sight.

These tiny displacements of the stars, or proper motions, were first detected by Halley in 1718; by comparing 17th-century observations with those of antiquity he showed that the positions of Aldebaran and Arcturus had changed, relative to the stars near them on the star sphere. Since that time, thanks to the improvement of astronomical instruments and methods of measurement, the number of proper motions (some small, some relatively large) that have been detected and measured has grown steadily. Today it stands at about 40,000.

The largest proper motion so far discovered is that of a magnitude 10 star in Ophiuchus, spotted by the American astronomer Barnard. Each year it moves across the star sphere a distance of 10.3″, which is comparatively so enormous an annual displacement that it has earned it the nickname of 'Barnard's arrow'. To make this displacement a little more meaningful — it amounts in 180 years to the diameter of the full Moon.

Nearly as large (8.8″) is the annual proper motion of another faint star, of magnitude 9, known as Kepteyn's star. After this, in order, come Groombridge 1830, a star of magnitude 6.4. annual proper motion 7.05″; Lacaille 9352 (magnitude 7.4), 6.9″; and a dozen others, including van Maanen's star (magnitude 12.3) with an annual displacement of 2.98″. Once a star's distance is known, the linear value of its proper motion in m.p.s. is easily calculated from its angular displacement. With a few exceptions — such as 61 Cygni ε Indi, o_2 Eridani, α Centauri, μ Cassioepiae and ε Eridani — the vast majority of stars with large proper motions are invisible to the naked eye. The motions that ought to be the most noticeable thus evade ordinary observation and, though they may have resulted in large displacements since the earliest descriptions of the heavens were made, they produce no alteration in the naked-eye appearance of the constellations.

The brighter star groups are also changing slowly, but for these to become noticeable we should have to travel back to prehistoric times, or equally far forward into the future. The appearance of the Great Bear as presented to the cave dwellers, for example, was different from that with which we are familiar; and in another 100,000 years it will again be unrecognisable (see Fig. 561). These reconstructions of the past and future outlines of a constellation can only be made when we know, with sufficient accuracy,

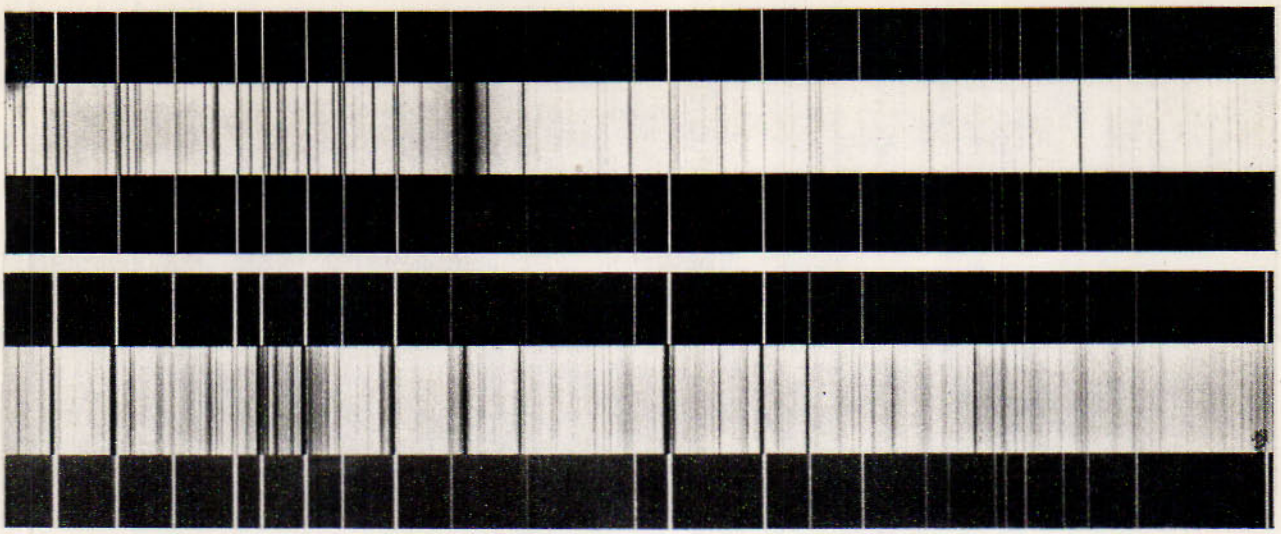


Fig. 563 — THE SMALL DISPLACEMENTS OF THE SPECTRAL LINES OF STARS, COMPARED WITH THE CORRESPONDING LINES OF A TERRESTRIAL SOURCE (IRON ARC) REVEAL THEIR RADIAL VELOCITIES OF APPROACH (—) OR RECESSION (+). Above: α Carinae, type FO, radial velocity +13 m.p.s. Below: α_1 Centauri, type GO_2, radial velocity −15 m.p.s. *(Photos, Lick Observatory)*

both the size and the direction of the proper motion of each of its stars.

So far we have only been considering changes of position that take place across the surface of the star sphere. Such displacements do not completely represent the stars' true motions in space, which occur in every conceivable direction relative to the Earth: across the line of sight, more or less obliquely to it, or directly approaching or receding. In the two latter cases the star will have no proper motion, since the whole of its spatial displacement takes place along the line of sight. Fig. 564 explains these various alternatives diagrammatically.

A star's proper motion is thus only a partial reflection of its real spatial motion, which cannot be discovered unless its direction is known — or, more precisely, unless its projection on the line of sight can be evaluated (its projection on the plane perpendicular to this — its proper motion — being already known).

Now there exists a technique whereby it can be immediately discovered whether a star's distance from the observer is changing, and if so in what sense — approach or recession. This is one of the many boons conferred by the spectroscope: as though its revelation of a star's chemical composition, temperature, dimensions and mass were not sufficient, it can also tell us the actual speed of the star's motion in the line of sight. The basis of this determination is the shift of the lines of a spectrum towards the red or the violet, according as to whether the source is receding from or approaching the observer — the Doppler effect which is discussed in Chapter 20. The shift of a star's spectrum can be measured by comparing it with that of some laboratory source which is stationary relative to the spectroscope; the amount of the shift is directly proportional to the star's line-of-sight velocity.

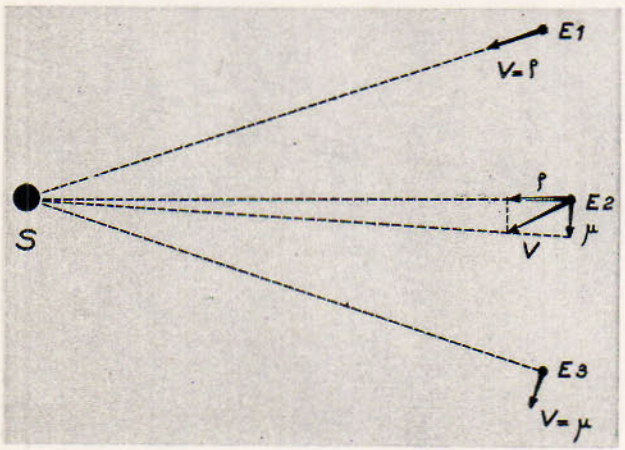


Fig. 564 — RESOLUTION OF V, THE ACTUAL MOTION OF A STAR RELATIVE TO THE SUN, into radial velocity (ϱ) and proper motion (μ), in different cases.

The component of a star's velocity which is revealed in this way is known as its radial velocity. It corresponds exactly to the space velocity of a star which is moving directly towards or away from the observer; in the case of a star moving obliquely across the line of sight, it is the component of that motion which is projected on the line of sight; and in the case of a star moving at right angles to the line of sight its value is zero (Fig. 564).

Once both the proper motion and the radial velocity of a star are known, the direction and velocity of its true motion in space can be derived.

Measured radial velocities are to a variable extent falsified by the motion of the Earth about the Sun; the Earth's orbital velocity (18 m.p.s.) must be allowed for, in order to determine the star's radial velocity relative to the Sun, which is assumed motionless. The majority of stellar radial velocities lie between about 10 and 40 m.p.s., either of approach or recession: some, however, far surpass these values, reaching 50, 100, or even 150 m.p.s. But, as will be realised from the foregoing discussion, their space velocities are almost always larger than their radial velocities. The record is held at the present time by a pair of stars numbered 5583 and 5584 in the Washington Catalogue, whose radial velocity of approach is 180 m.p.s., and total space velocity 360 m.p.s.

One could hardly suppose that the Sun alone among the stars is stationary in space — if, indeed, such a concept has any meaning — and the study of stellar motions as a whole has not only revealed the fact of the Sun's motion but has allowed this motion to be measured with some precision. It can only be recognised and measured by referring it to certain comparison points which are assumed, at any rate provisionally, to be fixed. At first, during the 18th and 19th centuries, the whole group of the nearer stars (whose proper motions, and later, radial velocities, were known) was made to provide reference

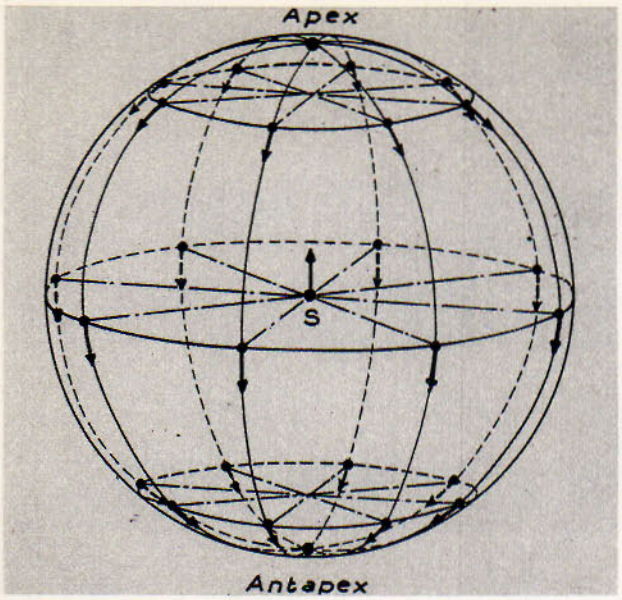


Fig. 565 — APPARENT PROPER MOTIONS OF THE STARS TOWARDS THE ANTAPEX, resulting from the Sun's own proper motion towards the Apex.

points. It is obvious that individual stars cannot be used, since each star has its own particular motion; only a group, within which the random individual motions of the stars will tend to cancel one another out, can be used as a reference system.

Granted this, the Sun's motion can be demonstrated without difficulty, either by the proper motions or the radial velocities of the stars in the group. It will readily be seen that if the Sun is moving in a given direction, the stars constituting the reference system will, by and large, appear to move in the opposite direction: this effect of relative motion is familiar to anyone who has watched the landscape sliding 'backwards' past the window of a railway carriage. Thus if we could watch the stars for several hundred thousand years we should notice them apparently approaching us from the direction of the Sun's motion and receding behind us in the opposite direction. Taking into account the effect of perspective, the proper motions of the stars would in general (Fig. 565) appear to be carrying them away from the point on the star sphere towards which the Sun is moving (the so-called Apex), while they would appear to be converging on the opposite point, the Antapex; the mean proper motions of stars at the Apex and Antapex themselves would be nil, and in all directions at right angles to the direction of the Sun's motion they would be maximal. In the same way, it can be seen that the radial velocities — of approach at the Apex, and of recession at the Antapex — would be maximal in these two directions and nil in those regions where the proper motions are maximal.

In this way it has been established that the Sun, relative to the *ensemble* of the nearer stars, is moving at a velocity of about 12. m.p.s. towards a point on the star sphere situated in the constellation Lyra not far from the first magnitude star Vega.

Later on we shall see how the study of the motions of ever more distant stars has revealed the existence of much more wholesale mass-migrations within our stellar system (see Chapter 16).

Although the appearance of the constellation figures is, generally speaking, an effect of perspective — making stars which are in fact very unevenly distributed in depth appear to be near to one another on the 'surface' of the star sphere — it may nevertheless be asked whether there is not some connection between the members of certain groups of stars: so that the whole or part of a constellation may in fact form a real cluster. The study of proper motions and radial velocities has provided an answer to this question.

Usually it is found that the motions of the different stars of a constellation figure are oriented quite at random — confirming our conclusion that their apparent mutual proximity is simply an effect of perspective. But there are certain exceptions to this rule: occasionally, velocities of the same order of magnitude, and oriented in more or less parallel directions, are observed. Such stars, without being 'near' to one antother in the ordinary sense, nevertheless form a physically connected unit and are voyaging through space together. They are said to belong to the same star stream, or to form a moving cluster.

Five of the principal stars of the Great Bear form such a moving cluster. The same thing is encountered among the stars of Orion, and with the two clusters of stars in Taurus known as the Hyades and the Pleiades.

We shall return later to the subject of star groups and star streams, and their significance.

Distances of the stars. — The distances which separate the stars from our terrestrial observatories — whose magnitude has already been emphasised in general terms — for a very long time defeated all attempts at measurement. In default of any definite evidence, astronomers had to content themselves with the conclusion that the stars are at least enormously distant. How, in the end, were these fantastically large distances measured? The detailed answer to this question must be deferred till Book IV, where we shall discuss the varied instrumental resources and observational techniques that have been applied to the study of the stars. But because of the fundamental importance of stellar distances we must at this point give some general idea of the principles and methods that have enabled us, first to measure the distances of the nearer stars, and then to extend the determinations to more and more remote ones. Such an account is also indispensable for the understanding of certain terms, current in stellar astronomy, whose use it would be impossible to avoid. Finally, it will enable us to understand why the precision of stellar distance determinations

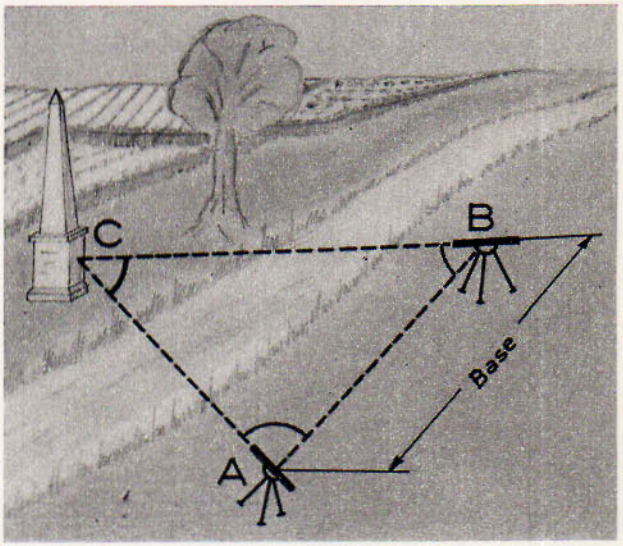


Fig. 566 — PRINCIPLE OF THE DETERMINATION OF THE DISTANCE OF AN INACCESSIBLE POINT BY MEANS OF ITS PARALLAX

falls far short of that with which the distances of the planets are known.

In theory, the measurement of the distance of an inaccessible point is an elementary problem in geometry. From two points A and B (Fig. 566), whose distance apart (the baseline AB) is known, the angles ABC and BAC are measured; in the triangle ABC we now know one side and all three angles, and can therefore calculate either of the other two sides, AC or BC — i.e. the distance of the point C. The apex angle, ACB, is known as the parallax of C.

Though this procedure is quite straightforward when it concerns an operation in terrestrial surveying, it becomes a matter of the utmost delicacy when applied to the rigorous determination of stellar distances. The reason for this will be evident enough. On the ground we can choose a baseline of sufficient length, compared to the other sides of the triangle, to ensure that the base angles are large enough for precise measurement. But when the object whose distance we are measuring lies outside the Earth the case is very different, for a new limitation is imposed on us — namely, that the longest baseline available on the Earth is its own diameter.

Despite this, the Moon is near enough to have an easily measured parallax — exceeding, in fact, 1°. For this reason its distance has been known very accurately for many years. In the case of the Sun, however, the diameter of the Earth is already a vanishing quantity, so that the solar parallax is so small an angle that it is difficult to measure accurately. And when we come to the stars, the triangle formed by the two ends of the Earth's diameter and even the nearest star is so elongated as to be indistinguishable from a straight line.

Fortunately, a way has been found out of this impasse. The size of the Earth may indeed be negligible compared with the distances of the stars, but the Earth is travelling round the Sun at a distance of 93 million miles; consequently, when an observer is carried from T_1 to T_2 (Fig. 567), which define the two ends of a diameter of the Earth's orbit, he is displaced in space a distance of 186 million miles (roughly 23,000 times the length of the Earth's diameter). This represents a much more impressive baseline — and yet its span is only just great enough to produce a few measurable stellar parallaxes; in the vast majority of cases it is still negligible. The diameter of the Earth's orbit, as seen from the *nearest* star, in fact subtends an angle of less than 2″ (2 seconds of arc), equivalent to 1 inch seen from a distance of over 1½ miles.

To simplify the foregoing account, parallax was defined as the angle subtended at the distant object by the baseline — the diameter of the Earth or of its orbit, as the case may be. In astronomical usage, however, it is taken to mean the angle subtended by half the baseline — i.e. by the radius of the Earth or its orbit. Thus the parallax of the Moon, planets and Sun refers to the semidiameter of the Earth, that of the stars to the semidiameter of its orbit.

No stellar parallax known at present exceeds 1″. This is equivalent to a distance of 206,265 times the radius of the Earth's orbit, a distance which is termed a parsec (*par*allax of one *sec*ond) and is the generally adopted unit for expressing stellar distances. The value of 1 parsec in miles is just short of 200 million million.

The advantage of employing the parsec as a unit of measurement is that it obviates the enormously large numbers that the use of miles would entail. Thus the distance of the nearest star is 1.31 parsecs — a much more convenient number to handle than 25,000,000,000,000 miles. The distances that are encountered once we leave the Earth are so enormous, however, that multiples of the parsec have had to be introduced: thus the kiloparsec and megaparsec are respectively 10^3 and 10^6 parsecs. Another frequently employed unit — and one that we shall use for preference since it conveys a more concrete notion of the distance involved — is the lightyear; this is the distance that light travels in one year. Its value in miles is roughly 5,860,000,000,000 or in parsecs 0.307; conversely, 1 parsec is equivalent to 3.258 lightyears.

The distances of our nearest stellar neighbours can be determined by the trigonometrical method, but that is not to say that all the comparatively near stars have been identified and their distances measured. Indeed, since the determinations are both prolonged and difficult, it is impossible to subject every star in the sky to the type of investigation that would eventually reveal its proximity — or otherwise. Astronomers have therefore focused their attention on stars for which there is some reason — in particular, large proper motion — to believe that they are not too remote. Very many have certainly escaped detection, owing to their faintness; many, indeed, are no doubt practically invisible.

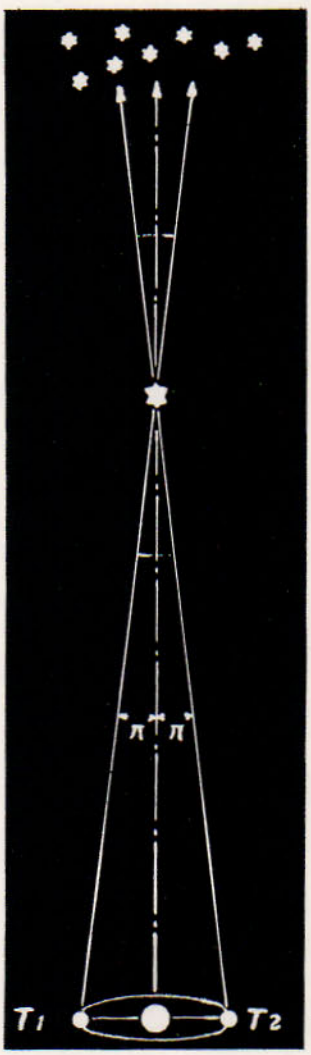


Fig. 567 — PARALLACTIC DISPLACEMENT OF A NEAR STAR, RELATIVE TO THE BACKGROUND OF MORE DISTANT STARS. (π = parallax; T_1, T_2 are the positions of the Earth at 6-month intervals.)

THE BRIGHTEST STARS IN THE SKY

STAR	APPARENT VISUAL MAGNITUDE	SPECTROSCOPIC TYPE	ANNUAL PROPER MOTION (″)	RADIAL VELOCITY (M. P. S.)	DISTANCE (L. Y.)	ABSOLUTE VISUAL MAGNITUDE	LUMINOSITY (⊙ = 1)
1. α Canis Majoris (Sirius)	−1.6	A0	1.32	− 5	8.6	+1.3	30
2. α Carinae (Canopus)	−0.9	F0	0.02	+12	100	−3.2	1,900
3. α Centauri A	0.1	G4, K1	3.68	−14	4.3	+4.7, +6.1	1.3
B							0.36
4. α Lyrae (Vega)	0.1	A0	0.35	− 8	27	+0.5	60
5. α Aurigae (Capella)	0.2	G0	0.44	+18	42	−0.4	150
6. α Bootis (Arcturus)	0.2	K0	2.28	− 3	33	+0.2	80
7. β Orionis (Rigel)	0.3	B8p	0.00	+15	540	−5.8	21,000
8. α Canis Minoris (Procyon)	0.5	F3	1.25	− 2	11	+2.9	6.9
9. α Eridani (Achernar)	0.6	B5	0.09	+12	70	−1.1	280
10. β Centauri	0.9	B1	0.04	− 7	190	−2.9	1,400
11. α Aquilae (Altair)	0.9	A5	0.66	−16	15.7	−2.5	10
12. α Orionis (Betelgeuse)	var	M2	0.03	+13	300	(−4)	(40,000)
13. α Crucis	1.1	B1	0.05	− 7	220	−2.7	1,200
14. α Tauri (Aldebaran)	1.1	K5	0.20	+34	53	+0.1	90
15. β Geminorum (Pollux)	1.2	K0	0.62	+ 2	29	+1.5	25
16. α Virginis (Spica)	1.2	B2	0.05	+ 1	120	−1.6	440
17. α Scorpii (Antares)	1.2	M1	0.03	− 2	250	−3.2	1,900
18. α Piscis Australis (Fomalhaut)	1.3	A3	0.36	+ 4	16	+2.0	23
19. α Cygni (Deneb)	1.3	A2p	0.00	− 2	400	−4.2	4,800
20. α Leonis (Regulus)	1.3	B8	0.25	+ 4	67	−0.3	130

THE NEAREST STARS TO THE SUN (after P. van de Kamp, 1945)

STAR	APPARENT VISUAL MAGNITUDE	SPECTROSCOPIC TYPE	ANNUAL PROPER MOTION (″)	RADIAL VELOCITY (M. P. S.)	DISTANCE (L. Y.)	ABSOLUTE VISUAL MAGNITUDE	LUMINOSITY (⊙ = 1)	REMARKS
1. Sun	−25.8	G0				+ 5	1.0	
2. α Centauri A	0.3	G4	3.68	− 14	4.3	4.7	1.3	
B	1.7	K1				6.1	0.36	
C	11	M?				15.4	0.00007	2° from A-B
3. Barnard's star	9.7	M5	10.30	− 70	6.1	13.3	0.0008	double
4. Lalande 21185	7.6	M2	4.78	− 54	7.9	10.7	0.0052	double
5. Wolf 359	13.5	M5e	4.84	− 56	8.0	16.6	0.00023	
6. Sirius A	− 1.6	A0	1.32	− 5	8.6	1.3	30	
B	(7.1)	A5				(10.0)	(0.010)	? double ? Mag.
7. Ross 154	11	M4e	0.67	?	9.3	13.7	0.00033	
8. Ross 248	12.2	M6	1.58	?	10.3	14.7	0.00013	
9. Luyten 789-6	12.3	M5e	3.27	?	10.3	14.8	0.00012	
10. *E* Eridani	3.8	K0	0.97	+ 9	10.7	6.2	0.33	
11. Procyon A	0.5	F3	1.25	− 2	11.0	2.9	6.9	
B	10.8	?				13.2	0.00052	
12. 61 Cygni A	5.6	K5	5.22	− 40	11.1	7.9	0.069	
B	6.3	K6				8.6	0.036	
13. Ross 128	11.1	M5	1.40	?	11.2	13.4	0.00044	
14. *E* Indi	4.7	K5	4.67	− 25	11.2	7.0	0.16	
15. τ Ceti	3.6	K0	1.92	− 10	11.2	5.9	0.44	
16. Strube 2398 A	8.9	M4	2.29	0	11.3	11.2	0.0033	
B	9.7	M5				12.0	0.0016	
17. BD −12° 4523	9.7	M4	1.24	?	11.6	11.9	0.0017	
18. Groombridge 34 A	8.1	M1	2.91	+ 5	11.7	10.3	0.0076	
B	10.9	M6				13.1	0.00058	
19. Lacaille 9352	7.4	M2	6.87	+ 6	12.0	9.6	0.014	
20. BD +5° 1668	10.1	M4	3.73	+ 15	12.4	12.2	0.0013	double
21. Ross 614	11	M4e	0.97	?	12.5	13.1	0.00058	double
22. Lacaille 8760	6.6	M1	3.46	+ 15	12.5	8.7	0.033	
23. Krüger 60 A	9.8	M4	0.87	− 15	12.7	11.8	0.0019	
B	11.3	M6				13.8	0.00048	
24. Kapteyn's star	8.8	M0	8.79	+150	12.7	10.8	0.0048	
25. van Maanen's star	12.3	F0	2.98	+148	13.2	14.3	0.00019	
26. Groombridge 1618	6.8	K6	1.45	− 17	14.1	8.6	0.036	
27. Wolf 424 A	12.6	(M5e)	1.87	?	14.2	14.4	0.00017	
B	12.6	(M5e)				14.4	0.00017	
28. CD −46° 11,540	9.4	M3	1.15	?	14.5	11.2	0.0033	
29. AOe 17 415-6	9.1	M4	1.31	− 10	15.1	10.8	0.0048	
30. Ross 780	10.3	M5	1.12	?	15.3	11.9	0.0017	
31. CD −44° 11,909	10.0	M5	1.14		15.4	11.6	0.0023	
32. BD +20° 2465	9.5	M3	0.49	+ 5	15.5	11.1	0.0036	double
33. CD −37° 15,492	8.3	M3	6.09	+ 15	15.5	9.9	0.011	
34. CD −49° 13,515	8.6	M3	0.78	?	15.6	10.2	0.0083	
35. Altair	0.9	A5	0.66	− 16	15.7	2.5	10	
36. BD +43° 4305	10.2	M5e	0.84	+ 1	15.7	11.8	0.0019	
37. o_2 Eridani A	4.5	G5	4.08	− 26	15.9	6.1	0.36	
B	9.2	B9				10.8	0.0048	
C	10.7	M4e				12.3	0.0012	
38. AC +79° 3888	11.0	M4	0.87	?	16.4	12.5	0.0010	
39. 70 Ophiuchi A	4.3	K1	1.13	− 4	16.5	5.8	0.48	
B	6.0	K5				7.5	0.10	

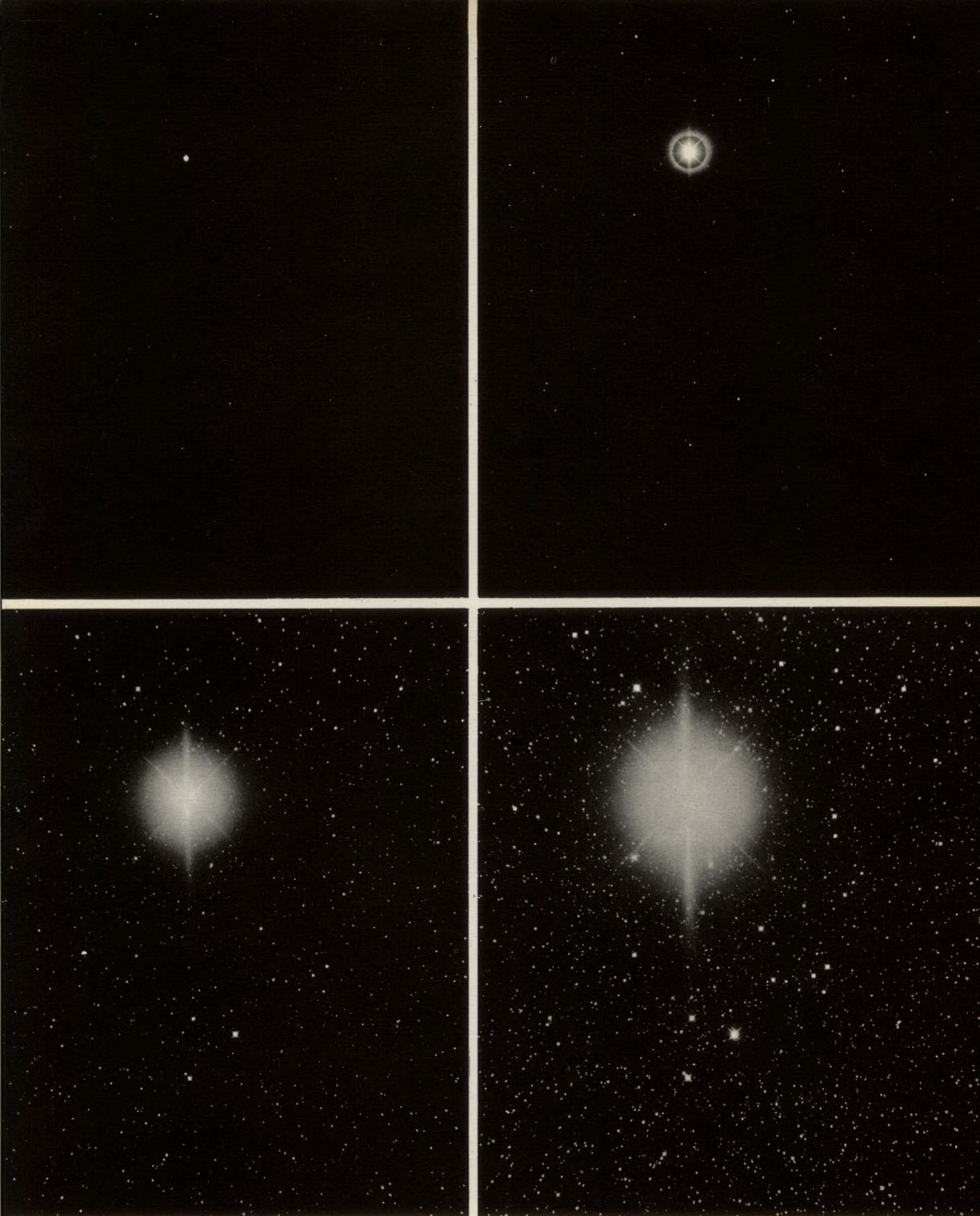


Plate VII. STARS OF DIFFERENT MAGNITUDES. The innumerable multitude of the stars becomes evident as fainter and fainter ones are recorded.

These four photographs, taken successively with instruments of increasing resolving power, show: upper left, stars to magnitude 12; right, to magnitude 15; lower left, to magnitude 18; right, to magnitude 20. The naked eye can see not a single star in this area of sky (i.e. none is brighter than magnitude 6). Note that the bright cruciform rays centred on the brighter stars are of purely instrumental origin.

Plate VIII. THE CONSTELLATIONS OF THE NORTHERN HEMISPHERE

(Courtesy, Geo. Philip & Son Ltd.)

Plate IX. THE CONSTELLATIONS OF THE SOUTHERN HEMISPHERE

(Courtesy, Geo. Philip & Son Ltd.)

Plate X. THE CONSTELLATIONS OF THE EQUATORIAL ZONE

(Courtesy, Geo. Philip & Son Ltd. and E. O. Tancock.)

The first stellar parallax to be measured by this (or any other) method — known as trigonometrical parallax — was that of 61 Cygni, a star whose large proper motion had suggested that it might be a relatively near neighbour. Its parallax was measured by the German astronomer Bessel in 1838, and at long last allowed astronomers to form a rough idea of the scale of stellar distances. It was, in fact, only a rough idea, since Bessel's determination was not very accurate, and had to be modified by subsequent workers. The delicate nature of these distance determinations is well shown by the successive values that were obtained as instruments steadily improved and as the growing number of determinations allowed a mean to be taken, thus eliminating the inevitable uncertainties: although Bessel attributed a parallax of 0.37″ to 61 Cygni, the currently accepted value is only 0.30″, corresponding to about 10 lightyears.

The stars which are spoken of as being 'near' are those whose distances are less than about 5 parsecs (16 lightyears). The table on p. 310 lists all these that are known at the present time.

We have learnt that the nearest star is 4.3 lightyears away. But this is nothing compared with the distances of myriads of the stars distributed throughout the depths of space; and the direct measurement of parallactic angles, which is already difficult even in the case of the nearer stars, grows less and less accurate as the angles become progressively smaller, finally becoming altogether impracticable when the diameter of the Earth's orbit itself becomes negligible compared with the distances to be measured.

Hence this direct trigonometrical method of measuring parallaxes can only be employed up to a certain limit; this limit admittedly depends on the instrumental power available, but at the present time it is about 100 parsecs (300 lightyears).

The distances of all stars situated beyond this limit would thus remain unknown if the method of trigonometrical parallax were the only one at our disposal. Fortunately, however, entirely different methods, of great mathematical and physical ingenuity, have been developed. The basic principles of these must now be briefly described.

The first of these methods makes use of the proper motions of the stars, and of our Sun in particular. We have already seen how, as a result of the Sun's motion through space, the stars should suffer displacements which are the perspective reflection of this motion. But this particular effect is vitiated in the case of individual stars by their own proper motions, and for the perspective effect of the Sun's motion to become apparent we must consider a suffciently large collection of stars for the individual proper motions to vanish in the mean. It will readily be seen that the magnitude of the apparent displacement of such a group will be a function of its mean distance: to return to the simile of the observer at the window of a railway carriage, the displacement of the countryside and of the hills on the horizon is much slower than that of the telegraph poles in the foreground. Thus we can see that once the absolute velocity of the Sun has been established by mass observations of radial velocities, the mean annual proper motion of the group of stars will reveal the distance of this group. Distances so derived are known as hypothetical parallaxes. The method is applicable at distances where the trigonometrical method is completely useless, for if the displacements are too small for accurate measurement, one only has to wait — 20 years, 50 years, 100 years, perhaps — and eventually they will grow to measurable proportions. One of the principal objects of the *Carte du Ciel* project was the determination of large numbers of just such proper motions.

Another method, that of dynamic parallax, is only applicable to double stars (Chapter 13). It involves the assumption that the mass of each star of such a pair is of the same order as that of the Sun; in general, this is approximately true. Making this assumption, and applying the laws of gravitation to the observed motions of the two components, the linear radius of the orbit can be derived. By comparing this with the observed angular radius the distance is at once obtained. Owing to the great number of double stars in the sky, this method has provided much valuable information.

But the most fruitful method of all is that of spectroscopic parallax. The practical details of the method will be more fully described later: in principle it consists in comparing the observed apparent magnitude of a star with its absolute magnitude (see below) as deduced from certain characteristics of its spectrum — in particular, the intensities of its absorption lines. The methods of determining distance which we have already described have provided the distances of a sufficiently representative collection of stars for it to be established that the intrinsic luminosities of the stars (i.e. their absolute magnitudes) are related to their fundamental nature, which in turn is reflected in their spectra.

Conversely, the recognition of these characteristic features in the spectrum of a star permits its intrinsic luminosity to be inferred; the comparison of this with its apparent brightness at once gives its distance.

Of the 20 brightest stars in the sky, listed on p. 310, the majority have revealed their distances by the spectroscopic method.

This physical method of distance determination is by no means rigorous, however: in fact it carries an uncertainty of about 20% in the derived distances. But though it may be less accurate than the trigonometrical method, when this is applied to relatively nearby stars, its indisputable advantage lies in the fact that it permits the plumbing of space to far greater depths. The accuracy of the trigonometrical method falls off with increasing distance even over the comparatively restricted distance range where it can be used at all. The accuracy of the spectroscopic method, on the other hand, is not affected by distance, and is equally applicable at all ranges, subject only to the one condition that the star should be bright enough for the details of its spectrum to be discernible. Since it is preferable to obtain — and in great numbers — slightly imprecise values than to obtain none at all, the importance of this method (initiated by the American astronomer Adams, about 40 years ago) will readily be appreciated.

The distances of some thousands of stars, obtained by the spectroscopic method, have already enabled us to survey an extensive region of the stellar domain — to some hundreds, and even thousands, of lightyears. In due course we shall see that other techniques have allowed us, little by little, to extend our distance determinations to the limits of the visible Universe.

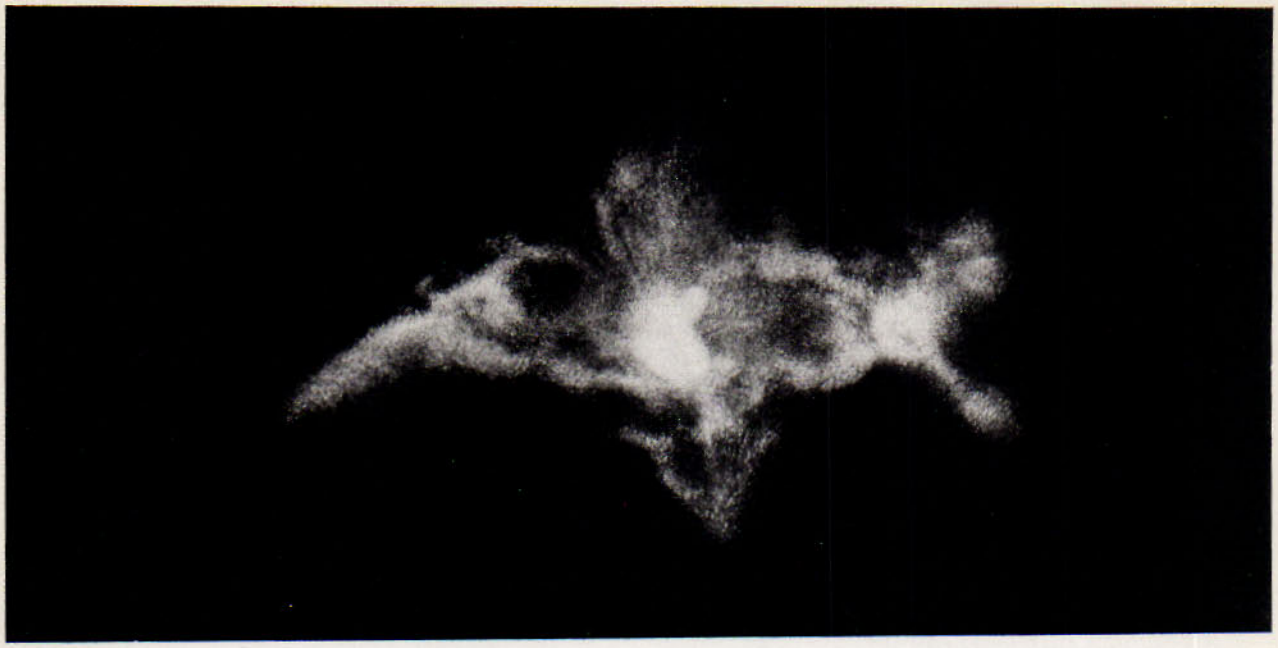


Fig. 568 — R AQUARII, A VARIABLE DOUBLE STAR SITUATED IN THE MIDST OF A STRANGE NEBULOSITY
(Photo, Mount Wilson Observatory, 100-inch telescope; exposure 8 hours)

CHAPTER 13

DOUBLE AND VARIABLE STARS

It will have been noticed in the foregoing tables that certain stars are designated 'double' or 'triple', or that the entry in the magnitude column reads 'variable'. Not all stars, in fact, are as solitary and as steady-shining as our Sun: many belong to associations of two, three, four or more members, while others again do not maintain a constant brightness. The study of these special, but very widespread, categories of star has received much attention in modern times, and has been productive of much valuable information. It is proper, therefore, that we should devote to them the space that their importance merits.

DOUBLE AND MULTIPLE STARS

To the naked eye, the stars are no more than isolated scintillating points. But telescopic observation reveals that many of them are in reality double or multiple — what appears as a single star to the naked eye is found to consist of two or several stars, lying more or less close to one another. In some cases the components are of the same, or very nearly the same, brightness; in others, the difference between them is very marked, a bright star being accompanied by one that may be barely visible.

Appearance and nature of double stars. — The occurrence of these stellar associations was originally thought to be an accidental effect of celestial perspective: the positions of the two stars in space being such that, although they may be very distant from each other, they lie near to the same line of sight from the observer. Such pairs are known as optical doubles; they certainly occur in large numbers over the whole expanse of the sky, but the systematic investigations of William Herschel, around the year 1780, and of many subsequent astronomers, demonstrated that as often as not there is a physical connection between the components of close pairs. By this is meant that the two stars are relatively near to one another in space and are effectively involved in each other's gravitational field, by virtue of which they are revolving about one another — a striking vindication, incidentally, of the universality of Newton's law. Optical

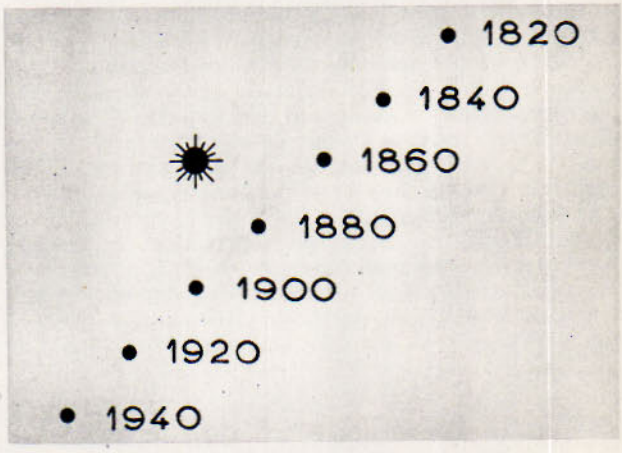


Fig. 569 — OPTICAL DOUBLES, DUE TO ACCIDENTS OF PERSPECTIVE, REVEAL THEMSELVES BY THEIR RECTILINEAR AND UNIFORM RELATIVE MOTIONS

doubles can therefore be clearly distinguished from these true binaries by the fact that the relative motions of their components are uniform and rectilinear, since their apparent association is both fortuitous and temporary.

Many binaries present a beautiful sight in the telescope owing to the strikingly contrasted colours of their components — a blue or green star (often described in the literature as 'sapphire' or 'emerald') may be accompanied by one of a bright yellow or orange tint, or even a clear red, resembling a glowing ember; in other cases the colours of the two stars are similar, though these always lie within the white and yellow range. The following table gives data of fifteen well-known binaries, whether coloured or not, all of which are observable with small telescopes.

STAR	MOST RECENT AVAILABLE DATA		MAGNITUDES AND COLOUR OF COMPONENTS
	SEPARATION (″ ARC)	DATE	
β Cygni	34.3	1955	3.2 orange — 5.4 blue
η Persei	28.0	1947	3.9 orange — 8.7 blue
α Canum Venaticorum	19.8	1947	2.9 yellow — 5.4 purplish
ζ Ursae Majoris	14.4	1955	2.4 white — 4.0 white
η Cassiopeiae	10.7**	1957	3.7 yellow — 7.4 red
γ Delphini	10.0	1955	4.5 yellow — 5.5 green
γ Andromedae	9.8	1955	2.3 orange — 5.1 green*
β Orionis	9.5	1947	0.3 white — 9.0 blue
γ Virginis	5.2**	1957	3.7 yellow — 3.7 yellow
α Herculis	4.7	1955	3.5 red — 5.4 green
γ Leonis	4.28	1957	2.6 yellow — 3.8 yellow
δ Serpentis	4.0	1955	4.2 bluish — 5.2 bluish
α Scorpii	2.9	1953	1.2 red — 7.0 green
ε Aurigae	2.8	1953	3.0 yellow — 6.3 blue
α Geminorum	2.25**	1957	2.7 white — 3.7 white

* The companion is itself double (5.5 green — 6.3 blue), with a separation of about 0″.5.
** Owing to the orbital motion of the *comes*, the separation of these stars varies comparatively rapidly.

Despite the undeniable beauty of some of these contrasted colours, it must be pointed out that in some cases the contrast is partly illusory — a physiological effect explained by the theory of complementary colours. The contrast between a yellowish and a bluish star thus tends to be exaggerated, while a white star (especially if faint) alongside a bright yellow or orange one will appear to be tinged with blue or green. Many binaries in reality lack the brilliant colours that have been attributed to them. Furthermore, a great diversity will be encountered among the colour-estimates of different observers; the 'greens' and 'violets', in particular, are wholly subjective.

Even if much of the colour contrast shown by the components of double stars is to be discounted as an illusion, however beautiful, it still remains true that their colours often are different, and that therefore the physical conditions of the components are different.

The study of binary systems is an extremely important branch of stellar astronomy, and today more than 25,000 binaries are known.

Visual and spectroscopic binaries. — Binaries constitute the most extensive class of multiple system. They range from wide pairs that a small telescope can easily resolve, down to those whose angular separation is so small that very powerful instruments are required in order to show the components clearly separated. The former may have separations of the order of 20″ to 30″; those of the latter do not exceed a few seconds of arc, or even one-tenth of a second of arc. Below this limit, not even the largest telescopes can separate them, and they remain to all appearances a single star.

Closer pairs nevertheless exist, though their existence would have escaped discovery but for the assistance, once again, of the spectroscope. We saw in the last chapter how the displacement of the lines of a spectrum by virtue of the line-of-sight motion of its source has been exploited. In 1889 the American astronomer E. C. Pickering detected a periodic doubling of the spectral lines of ζ Ursae Majoris (Mizar) — or, more precisely, of the brighter component of this star, which has long been recognised (Riccioli, 1650) as a visual double. Thus, by the application of Doppler's principle, the first spectroscopic binary was discovered: this star, apparently single in even the largest telescopes, consists in reality of two components which are revolving about one another at too small a distance to permit their visual separation. The study of their combined spectrum showed that their mutual revolution is accomplished in about 20 days at a mean distance of 102 million miles (slightly greater than that of the Earth from the Sun); this, at Mizar's distance of 80 light years, is reduced to an angular separation of 0.04″. Though this is much too small to be perceived with any existing visual instrument, it is nevertheless measurable with the interferometer (described in Chapter 19). Discoveries of

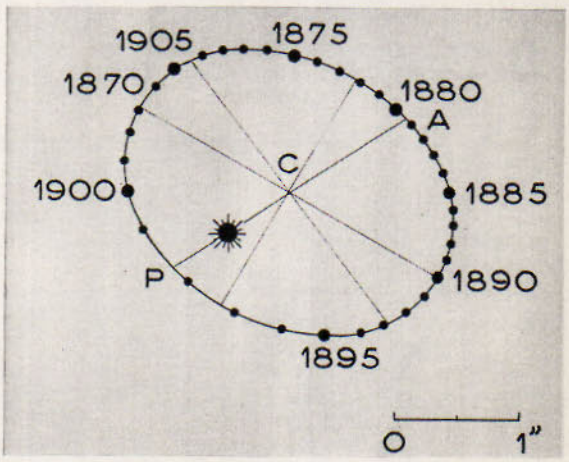


Fig. 570 — PHYSICAL DOUBLES (BINARIES) REVEAL THEIR TRUE NATURE BY THE ORBITAL MOTIONS OF THEIR COMPONENTS. The example illustrated here is ζ Herculis, period 35 years.

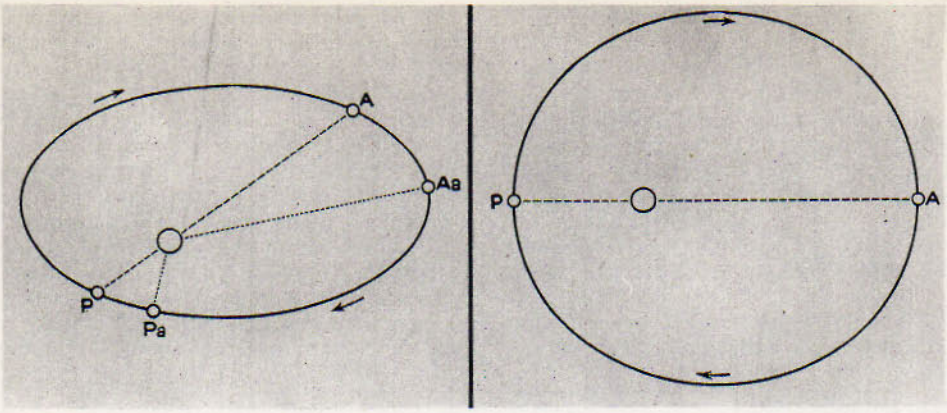


Fig. 571 — RELATION OF APPARENT ORBIT (LEFT) TO TRUE ORBIT (RIGHT) OF A DOUBLE STAR. A, true apoastron; Aa, apparent apoastron; P, true periastron; Pa, apparent periastron.

this kind followed thick and fast, and today more than 1,500 spectroscopic binaries are known.

The majority of multiple systems consist of a combination of these two types of binary, visual and spectroscopic; one component of a visual binary is found to be itself a spectroscopic pair, as in the case of Mizar, thus forming a triple system. If both components of a visual pair are spectroscopic binaries, a quadruple system results — Castor (α Geminorum) being an example. Besides these, visual triple systems are also encountered, the best known examples being ζ Cancri and ξ Scorpii: the two principal components are close to one another and revolve rapidly about their common centre of gravity, while the third member revolves more slowly, and at a greater distance, about the central pair. The relations between these motions are extremely complex, being in fact an example of the famous three-body problem that has taxed the ingenuity of the greatest mathematicians.

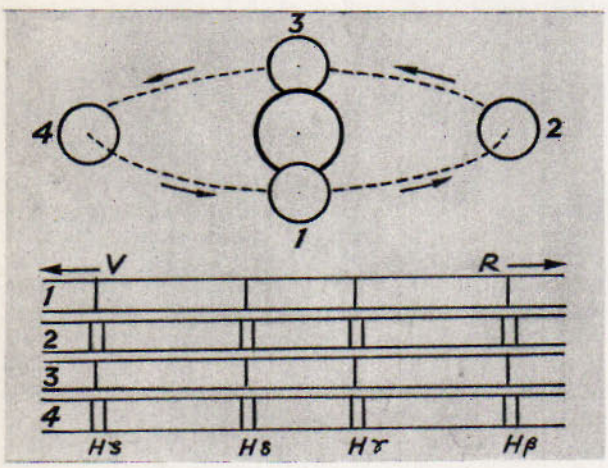


Fig. 572 — DIAGRAMMATIC EXPLANATION OF THE DOUBLING OF THE SPECTRAL LINES OF A SPECTROSCOPIC BINARY

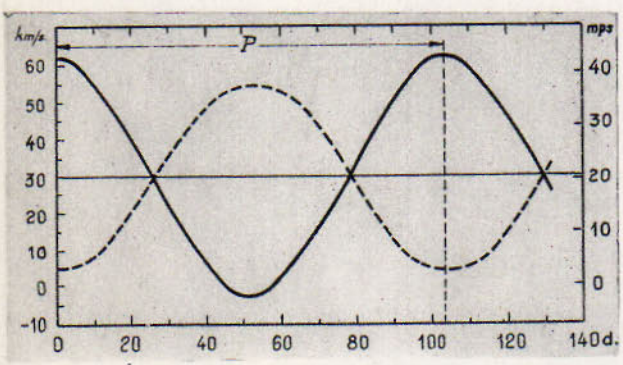


Fig. 573 — RADIAL VELOCITY CURVES OF THE TWO COMPONENTS OF THE SPECTROSCOPIC BINARY, α AURIGAE (CAPELLA)

The orbital motions of the components of visual binaries are often inappreciable, or almost so. Three-quarters of the doubles that have been studied during the past hundred years have shown no change of relative position, or at most have moved through an angle of a few degrees. Despite this, however, the true binary nature of a pair is established whenever the components are found to have a common proper motion. It is logical to suppose, in such cases, that the apparent absence of mutual revolution springs from the length of the period, perhaps numbering thousands of years; over a century or less, the displacement would be undetectable. On the other hand, some pairs revolve comparatively rapidly, and a few have even completed several revolutions since they were first observed carefully — i.e. in little more than a century.

The determination of the orbits in such cases is a comparatively simple matter, and validates the extension of Kepler's 'ellipse' law from planets to stars. It must be pointed out, however, that we are not now speaking of the directly observable apparent orbit but of the true orbit, whose spatial orientation is quite fortuitous and of which the apparent orbit is merely the projection on the star sphere — i.e. a perspective view, whose resemblance to the true orbit depends on the inclination of the latter to the line of sight. The two orbits are coincident only when, as sometimes happens, the plane of the true orbit is perpendicular to the line of sight (ζ Cancri is an example). It may also happen that the orbit is viewed almost exactly 'edge on', when the *comes** will appear to oscillate back and forth along a short, straight line (e.g. 42 Comae). In order to determine the elements of the true orbit, therefore, the apparent orbit must first be corrected by geometrical or analytical methods.

Lastly, it is important to realise that what we call the true orbit is the path the *comes* appears to trace out in circling its primary, which is assumed to remain motionless. In fact, we know that in a system of this kind both stars will be in orbital motion about their common centre of gravity; the two orbits will be of the same general shape, but their sizes will be inversely proportional to the mass of the star concerned. The centre of gravity may be

* i.e. the fainter, companion star. Plural: *comites*.

FEATURES OF SOME VISUAL AND SPECTROSCOPIC BINARIES (after P. Baize)

STAR	MAGNITUDES	PERIOD		ECCENTRICITY OF ORBIT	SEMI MAJOR AXIS	
		DAYS	YEARS		" ARC	10^8 MILES
W Ursae Majoris	7.9	0.33		0.00		0.95
α Geminorum C	9.0	0.81				1.60
V Puppis	4.1	1.45		0.08		5.48
β Aurigae	2.1	3.96		0.00		7.64
ζ_1 Ursae Majoris	2.4	20.54			0.04	26.4
α Aurigae	0.2	104.02	0.285	0.02	0.05	(52)
β Capricorni	3.2	1,375.3	3.77	0.44		(234)
δ Exuulei	5.4, 5.4		5.70	0.42	0.26	385
ε Hydrae AB	4.1, 4,4		15.04	0.61	0.21	1,030
42 Comae	5.2, 5.2		25.87	0.52	0.66	1,140
ζ Herculis	3.1, 5.9		34.38	0.48	1.37	1,270
ξ Scorpii AB	4.9, 5.0		45.69	0.74	0.72	1,860
Sirius	–1.5, 8.4		49.94	0.59	7.62	1,900
ζ Cancri AB	5.6, 6.1		59.60	0.31	0.95	1,960
α Centauri	0.4, 1.6		80.09	0.52	17.66	2,200
70 Ophiuchi	4.3, 6.0		87.85	0.50	4.56	2,200
φ Ursae Majoris	5.2, 5.3		108.9	0.47	0.32	4,300
γ Virginis	3.7, 3.7		171.4	0.88	3.75	4,200
η Cassiopeiae	3.7, 7.4		400.9	0.43	10.95	6,000

regarded as a fixed point in the system, though it may of course have a simple motion of translation relative to the other stars in the Universe. We have already had occasion to describe such a combination of motions when dealing with the Earth-Moon system (Chapter 4).

Among the 25,000 or more visual binaries now known, there are not more than 200 whose orbits have been determined satisfactorily, but even this small number is large enough for us to deduce certain important general characteristics. We should also add the 300-odd orbits of spectroscopic binaries that have been determined up to the present time, which lack of space prevents us from discussing separately. In any case the distinction between the two categories of binary depends primarily on accidental factors – the linear separation of the components, and their distance from the observer. If a pair is both widely separated and fairly near to the Sun, it will be a visual binary; if its separation is too small or its distance too great, its resolution will not be possible visually, and we shall have to rely on the spectroscope. This in turn presupposes (*a*) that it is bright enough for the details of its spectrum to be made out, which with existing equipment becomes difficult at about the sixth or seventh magnitude, and (*b*) that its orbit is suitably oriented for the spectroscopic detection of its components' motions.

It may be noted, in passing, that our census of spectroscopic binaries is for this reason still very incomplete, being limited practically to those that are nearer than, at the most, a few hundred light years from the Sun.

Nevertheless, one can get a good idea of the frequency of binary and multiple stars from the table of the nearest stars, on p. 310, where it will be seen that of 51 stars, 16 are members of binary systems, and another 6 of triple systems. It is, indeed, estimated that at least 1 star in 3 – possibly even 1 in 2 – is a member of some sort of multiple system. In any event, such stellar associations cannot be regarded as rare freaks: on the contrary, we are face to face with an extremely general characteristic of the stellar universe.

Returning to the subject of binary orbits, let us first consider the principal features of a number of visual and spectroscopic orbits that have been well determined. (See table above). It will be noticed that revolution periods vary within very wide limits – from a few hours to several months or years in the case of spectroscopic binaries, and from a few years to several centuries (and probably very much more) in the case of visual binaries. We have seen that the distinction between the two categories is artificial in that it is determined only by the conditions of their observation; certain binaries can in fact be observed both visually and spectroscopically, and the elements furnished by the two methods are in good agreement.

It will also be seen that the orbital eccentricities, which are for the most part small in the case of spectroscopic binaries, become more marked for those of moderate periods, while very long periods are commonly associated with highly eccentric orbits. It is permissible to wonder whether this correlation between period and eccentricity may not also be simply a selective effect of the conditions of observation. It appears, however, that this is not so; this rather puzzling correlation is probably connected in some way with the mechanism of the formation of binary systems.

Another point to notice is that the separations of fairly long period binaries are quite considerable, being comparable with those of the major planets from the Sun – or, rather, with those of the long-period comets, whose highly eccentric orbits are much more akin to those of visual binaries than the nearly circular orbits of the planets. The separations of spectroscopic components, on the other hand, are very much smaller, and compare with the solar distances of the terrestrial planets.

It may even happen that the separation of the two components is so small that they revolve almost in contact with one another. If the orbit is suitably oriented to our line of sight we shall observe the remarkable phenomenon

CORRELATION BETWEEN PERIOD AND ECCENTRICITY IN THE ORBITS OF DOUBLE STARS (after G. Aitken)

SPECTROSCOPIC BINARIES		VISUAL BINARIES	
PERIOD (DAYS)	ECCENTRICITY	PERIOD (YEARS)	ECCENTRICITY
2.7	0.05	16.8	0.43
7.6	0.16	37.1	0.40
14.1	0.22	73.0	0.53
30.6	0.35	138.0	0.57
102.5	0.30	274.3	0.62
1,177	0.31	2,000 *approx*	0.61
		5,000 *approx*	0.76

of two stars alternately eclipsing one another – a topic to which we shall return later.

But first we must examine an interesting problem, which has long been discussed and to which the study of binary systems has now provided at least a provisional answer.

Invisible stars and planetary comites. – As soon as it was suspected that the stars are merely suns – and still more when this became firmly established – the question arose whether they too might not be the centres of planetary systems, on the pattern of our own Sun.

Let it be said at the outset that not even the most powerful telescope in existence could reveal any planets, even giant ones, that might be revolving around these incredibly remote suns. It must be remembered that the stars, despite their great size, are visible only by virtue of their brilliance: without exception, they are mere dimensionless points of light. Thus it is evident that even if they were accompanied by planetary bodies, much smaller and infinitely less bright (for the planets, after all, are not efficient reflectors of the Sun's light), these would be far beyond the visual reach of the terrestrial observer.

However, we must not conclude from this that the problem is insoluble. In the first place, future instrumental and technical improvements may confidently be expected to reveal many things that are now hidden. But in the meanwhile, is there no indirect method of uncovering the existence of these hypothetical bodies? The answer is affirmative: observation has in fact confirmed this conjecture, based solely on the argument that since the Sun is a star, and has planets, it is at least possible that other stars are similarly accompanied.

We have said that both components of a binary revolve about the centre of gravity of the whole system. In the same way the Moon imposes some motion upon the Earth, which consequently traces out a small orbit about the Earth-Moon centre of gravity at the same time as the latter is revolving about the Sun in the Keplerian orbit.

With this principle as a starting point, it is easy to see that if a star were accompanied by sufficiently massive (though invisible) planets, their presence might be revealed through small periodic variations of the star's position: it would in fact appear to oscillate about a mean position. The angular value of such a displacement, and its irregularities, would give a clue to the motion and mass of the perturbing body or bodies, and the linear dimensions of the system could be derived as soon as its distance was known.

Let us see, now, how displacements of this kind – which are at best extremely minute – can be observed. All the stars are in motion, travelling through space in every conceivable direction relative to the Sun, and with velocities that may be large or small. These space motions are observed visually in the form of proper motions upon the star sphere. The proper motion of a single star is necessarily uniform, and its yearly positions will plot a straight (if very short) line. But in the case of a binary or multiple system, the perturbations of the *comites* will come into play. Only the centre of gravity of all the associated bodies will pursue a rectilinear course, while the positions of the components will alter as they revolve about it. The successive positions resulting from these circular or elliptical motions, when combined with the steady displacement of the system as a whole, will plot out a sinuous curve across the star sphere. Therefore, if we observe an apparently single star pursuing such a course, we know that it must be accompanied by one or more invisible bodies which, with it, are revolving round their common centre of gravity. Furthermore, the relative size of the system and the duration of its internal movements may be deduced from the observed periodicity of these departures from rectilinear motion.

A number of discoveries of this sort have already been made in the field of double star studies. The first, and most famous, was that of the companion of Sirius, during the 19th century. Irregularities in the proper motion of this magnificent star attracted the attention of Bessel, who in 1844 suggested that they were attributable to an invisible companion star; in 1851 Peters, basing his calculations on Sirius's observed sinuous proper motion, calculated the theoretical orbit of this body, which was still invisible in all the telescopes then available. Finally, in 1862, the American astronomer Alvan Clark discovered it visually

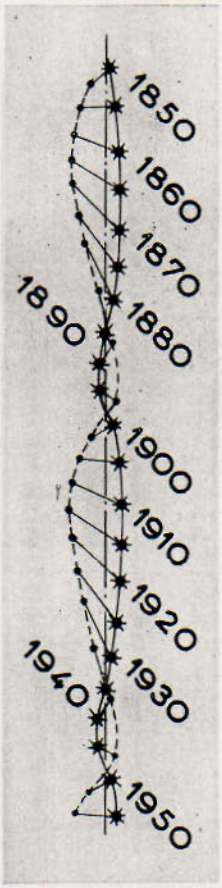


Fig. 574 – The sinuous proper motion of Sirius revealed the existence of its companion long before it had ever been seen.

in almost exactly the position that theory had predicted. This achievement (reminiscent of the discovery of Neptune) was repeated in 1896 with Schaeberle's discovery of the minute companion of Procyon with the great Lick refractor – likewise predicted by Bessel.

Since then, other discoveries of the same nature have been made, the non-rectilinear proper motion of an apparently single star showing that it is in fact a member of a binary system.

The method may be simple in theory, but its application is fraught with the greatest practical difficulties, owing to the extreme precision of the measurements necessary to reveal the almost imperceptible variations in the position of the star under observation. Where double

stars are concerned, the irregularities in the proper motion of the primary are often relatively noticeable, since the perturbations are caused by a body comparable in mass with the Sun. But the masses of planets are in general very much smaller, their effect upon a central sun correspondingly weaker, and the latter's departures from rectilinear motion infinitely less noticeable. For this reason it was for a long time impossible to detect the presence of planetary bodies outside our Solar System, and to this day it remains a feat of the greatest difficulty.

However, thanks to the precision of the photographic methods now applied to the study of double stars, positive results have been obtained in the case of at least two stars — 61 Cygni and 70 Ophiuchi, both of which have long been observed as visual binaries. The microscopic examination of photographs of both these binaries, taken at regular intervals over the preceding thirty years, enabled van de Kamp and Strand to announce in 1942 that systematic irregularities in the proper motions of both pairs were explicable on the assumption that they were accompanied by invisible third components. The visual system of 61 Cygni consists of two magnitude 6 stars; the revolution period is probably 720 years, and the major axis of the vast orbit subtends 25″. Since the measurements refer only to the relative positions of the two components, which are of sensibly equal brightness, we cannot say which of them is accompanied by the perturbing body. In either case the result would be the same: the unknown companion must move in an extremely eccentric orbit which brings it to within 0.7 A.U. of its primary. But the really interesting point is the extraordinarily low value of its derived mass — only 16 times that of Jupiter, or 10 times less than that of the lightest known star. For this reason alone it has been considered as a mammoth planet rather than a star.

The case of 70 Ophiuchi suggests a similar conclusion. Here the two members of the visual pair are unequally bright, the angular radius of the orbit is 5″, and the period of revolution is 88 years. As with 61 Cygni, it cannot be proved which star the invisible satellite is attached to. However, it has been possible to establish the following alternatives: if it revolves about the primary of the visual pair, the radius of its orbit must be 6.8 A.U. and its mass equal to 12 or 13 times that of Jupiter; if about the *comes*, the radius of its orbit is 5.9 A.U., and its mass 8 or 9 times Jupiter's.

Other 'super-planets' have since been suspected, though none of these is as well authenticated as 61 Cygni or 70 Ophiuchi. Ross 614 B, for example, has an estimated mass of about 0.08 ⊙ (roughly 85 times Jupiter); the combined mass of the two components of L 726—8 appears to be of the same order, making each star only about 40 times more massive than Jupiter. The two components of DQ Herculis, on the other hand, are both suspected of having masses as low as 0.006 ⊙, or only about 6 times Jupiter's.

The small masses of these invisible bodies — especially in the last case — justify our regarding them as planets (whose existence we should in any case expect, on *a priori* grounds), even though they are considerably, more massive than even the weightiest members of our Solar System.

According to this view, therefore, a great stride forward has been taken in our progressive discovery of the different types of object that make up the material universe; and we have every reason to believe that the increasing precision of the methods of astronomical mensuration will lead to further discoveries of the same kind. For it is clear that the cases we have just cited are in no way exceptional: 61 Cygni and 70 Ophiuchi are among the Sun's closest neighbours, and rather than suppose that

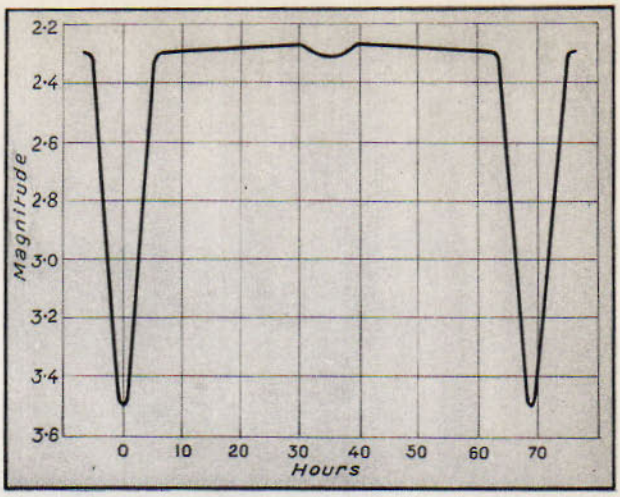


Fig. 575 — LIGHT CURVE OF AN ECLIPSING VARIABLE: β PERSEI (ALGOL).

it is mere chance that three stars in a comparatively small volume of space should have planetary systems, it is more logical to conclude that this very proximity is the reason for our being able to detect them. It is therefore permissible to assume that many other stars have planets revolving about them, and that their discovery only awaits the development of more powerful means of observation.

The term 'planetary system' may appear to be something of an exaggeration when, in the case of both the stars we have just considered, a single 'planet' is involved. But we must remember that other and less massive bodies revolving about these stars would cause no detectable perturbations. If we reverse the conditions of the problem and imagine that our Solar System is being studied by

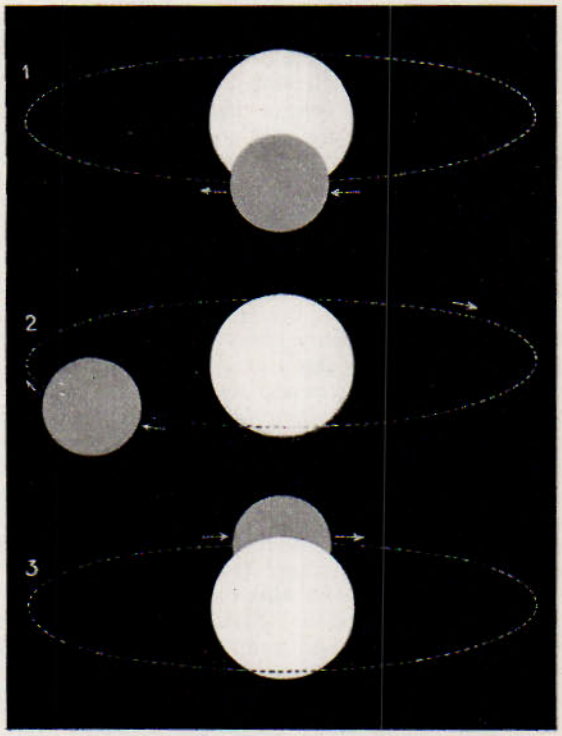


Fig. 576 — EXPLANATION OF THE LIGHT VARIATIONS OF ALGOLID ECLIPSING VARIABLES. 1, principal minimum; 2, outside eclipse; 3, secondary minimum.

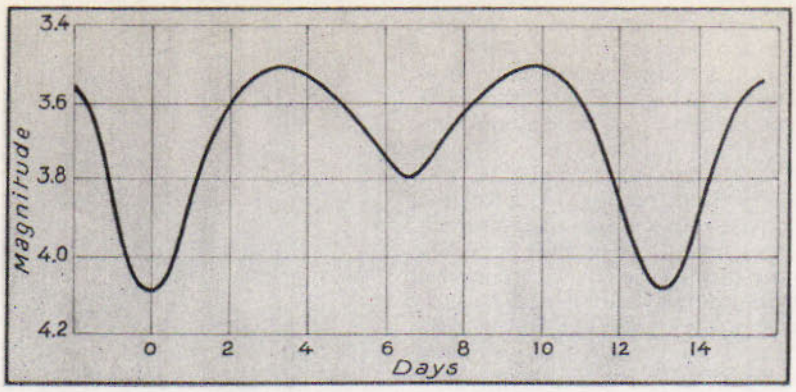


Fig. 577 — LIGHT CURVE OF THE ECLIPSING VARIABLE β LYRAE

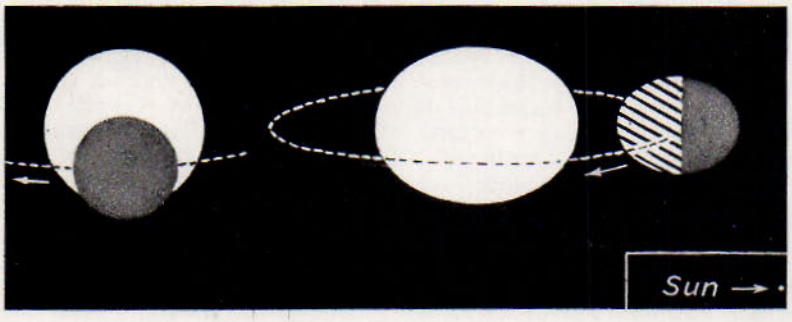


Fig. 578 — EXPLANATION OF THE LIGHT VARIATIONS OF ECLIPSING VARIABLES OF THE β LYRAE TYPE. Left: principal minimum; right: about maximum brightness.

an observer from a nearby star, who is employing the same method of investigating the Sun's planetary family, then of all the planets only Jupiter would have any appreciable effect on the Sun's motion — and even that would be so small as to be virtually negligible.

Thus the discovery of a single planet revolving round a star justifies us in assuming the existence of a whole 'solar' system — especially since, as we shall learn later, there is good reason to believe that planets are not born singly, but that (if they are born at all) they are produced in whole families.

VARIABLE STARS

These are stars whose brightness does not remain constant but varies as a function of time. The earliest to be studied scientifically was the spectacular 'new' star of 1572, which was carefully observed by Tycho Brahe. A quarter of a century later, in 1596, Fabricius noticed the strange disappearances and reappearances of the star *o* Ceti, which seemed to him to justify christening it Mira Ceti ('the wonderful star in Cetus') which name it has retained to this day. The fact that its variation is periodic escaped notice until 1638. Following this, the sudden decreases in the brightness of Algol (β Persei) were detected by Montanari in 1669, though their periodicity was not recognised till more than a century later, in 1782, by Goodricke, who also contributed the correct explanation of the phenomenon. Arab astronomers, however, may well have noticed it some centuries earlier when they gave the star its name — *al ghul*, the ghoul or demon.

It was nevertheless not until the middle 19th century that, at the instigation of Argelander, the systematic study of stellar variability was undertaken. And during the last half century it has developed so rapidly that today more than 20,000 variables are known and regularly observed.

Like other stars, variables are designated by a letter or letters, followed by the name of their constellation, but it has been agreed to reserve for them the capital letters of the Roman alphabet, starting with R for the first to be discovered in any given constellation, followed by S, T . . . Z; when this series is exhausted, the double letters RR, RS . . . RZ are used; then SS . . . SZ, TT . . . TZ, etc; after ZZ, the nomenclature goes on with AA, AB . . . AZ, followed by BB . . . BZ, etc., until QZ is reached. If, finally, the number of variables in any one constellation exceeds the number of these combinations (334), subsequent discoveries are denoted by V335, V336, etc. In addition to this system, bright stars retain the generally accepted designation that they already had (β Persei, β Lyrae, *o* Ceti, etc).

The observation of large numbers of variables has made it possible to classify them into several well-defined types, according to the mechanism of their variation.

Eclipsing binaries. — This class of variable (to which we have already referred in our discussion of binary stars) was the first whose mechanism was deciphered. They are not stars of intrinsically variable brightness, but simply binaries (more specifically, spectroscopic binaries) whose orbital planes happen to be parallel with the direction of the Earth; hence with each revolution the components eclipse one another, the eclipses being partial or total according to the relative sizes of the stars and the inclination of their orbit. By these mutual eclipses a very characteristic type of brightness-variation is produced: the combined brightness of the pair, observed with the naked eye or telescope as a single star, suffers sudden drops of fairly short duration, separated by equal periods of nearly constant brightness — the minima being the periods of eclipse.

The prototype of the eclipsing binaries is the beautiful β Persei (Algol), normally of about magnitude 2.2; at intervals of 69h it fades to about magnitude 3.4 (its brightness thus being reduced to one-third of its original value) in 5h, and immediately returns to its normal value in the same period of 5h. The way in which the brightness varies as a function of time can be represented by a regular curve (Fig. 575), known as its light curve. Thorough analysis of these curves — particularly when, as is the case with Algol, they can be supplemented by curves of the radial velocities of the stars, considered as spectroscopic binaries — allows many characteristics of an eclipsing system to be derived: the dimensions and shape of the orbit, and the sizes, shape, brightness and mass of both components. In this way it has been established that the components of Algol are relatively close to one another, and of roughly the same size (3 or 4 times that of the Sun); one is much hotter, and therefore brighter, than the other, and is partially eclipsed by it at every revolution. It will be seen that midway between the principal minima the light curve dips to a shallow secondary minimum, which has been detected by very precise photoelectric methods; this secondary minimum is caused by the partial eclipse of the companion star by the brighter component, proving that of the former is not completely dark but only very much fainter than the primary. In fact, measurements have shown that the surface of this 'dark' component is as bright as that of the Sun; we are reminded of the sunspots, which are actually more brilliant than an electric arc although they appear 'dark' by comparison with the photosphere.

It can also be seen from the light curve that the brightness of the system is not constant even when no eclipse is taking place, since the corresponding sections of the curve are not quite horizontal. This is explained by supposing that the much fainter *comes* re-emits some of the light that it receives from the primary: under such circumstances, we shall receive a varying fraction of this reflected light, according to whether the 'dark' hemisphere (at principal minimum) or the 'light' (at secondary minimum) is turned towards the Earth. Thus the over-all brightness of the system varies slightly outside eclipse, alternately increasing and decreasing during each successive half-period. When this secondary effect is absent, and the eclipses are moreover total, the light curve presents the perfect regularity of theory; such is the case with VV Cephei, whose investigation has revealed the extraordinary association of a hot and brilliant star revolving in a period of about 20 years about a gigantic cool star. We shall return to this later.

ELEMENTS OF SOME ECLIPSING VARIABLES
(after C. and S. Gaposhkin)

STAR	MAGNITUDE* MAX.	MAGNITUDE* MIN.	PERIOD (DAYS)	SPECTROSCOPIC TYPES OF THE COMPONENTS
W Ursae Majoris	7.9	8.5	0.33	F8 — F8
V Puppis	4.1	4.7	1.45	B1 — B2
u Herculis	*4.6*	*5.3*	2.05	B3 — B3
δ Librae	4.8	5.9	2.33	A0 —(G0)
β Persei	2.2	3.4	2.87	B8 —(G4)
λ Tauri	3.8	4.1	3.95	B3 —(K2)
β Aurigae	2.1	2.2	3.96	A1 — A1
δ Orionis	2.5	2.6	5.73	B1 —(B2)
β Lyrae	3.4	4.2	12.92	B5 — B8
α Coronae	*2.3*	*2.4*	17.36	A0 — ?
ζ Aurigae	*4.9*	*5.6*	972.15	K5 — B9
VV Cephei	4.9	5.7**	7,430 (20 years)	M2 — B9
ε Aurigae	3.1	3.9***	9,888 (27 years)	F5p — ?

* Photographic magnitudes in italics.
** Next eclipse in 1976; primary an irregular variable.
*** Next eclipse in 1982; the eclipse lasts two years, during one of which it is total.

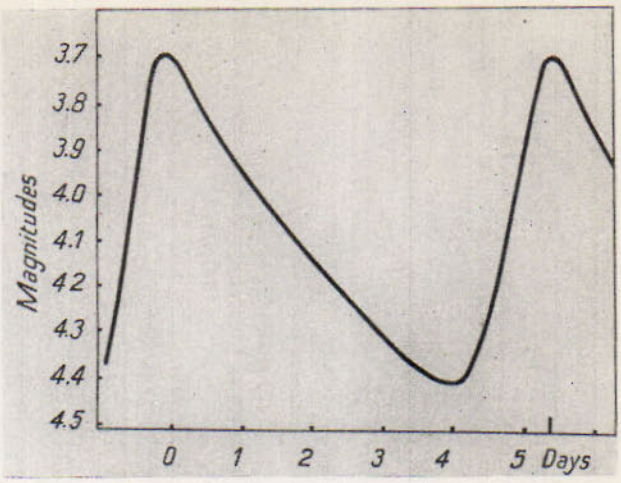


Fig. 579 — LIGHT CURVE OF δ CEPHEI

There is another secondary cause of the light variation of the eclipsing binaries, however, which results in a light curve which is at no time even approximately horizontal, but consists of two rounded maxima situated roughly midway between the minima. This is exemplified by the curve of β Lyrae (Fig. 577). This characteristic is attributed to the ellipsoidal forms of the two components; such systems consist of stars which are at the same time very large and very near to one another, under which circumstances their spherical shape is deformed into a marked oval by their mutual gravitation — rather as though it were a gigantic tidal effect. Furthermore, they always turn the same face towards one another (apart from a small libration effect), so that the area of their surfaces that is turned towards the Earth is constantly changing as the orientation of the major axes of the ellipsoids to the line of sight changes. The greatest area will evidently be seen (Fig. 578) at maximum elongations, roughly midway between eclipses.

This effect of stars distorted from the normal spheroid is sometimes so marked that it produces observable variations in the total brightness of spectroscopic binaries whose orbital plane is too highly inclined to the line of sight for eclipses to occur: these are the so-called ellipsoidal variables, of which ζ Andromedae and π Cassiopeiae are examples.

Other ellipsoidal eclipsing binaries of small stars, comparable with the Sun in size, are so close to one another that their revolution periods are frequently less than 12h and in some cases even drop below 5h. Systems of this kind (of which the type-star is W Ursae Majoris) are distinguished by their extremely rapid light variation and by the near-equality of the two minima, due to the almost twin-like similarity of their components.

Such are, in outline, the main features of the eclipsing binaries. They are much more than interesting or curious spectacles; they are important because their specific type of behaviour permits us to derive much valuable information about stellar sizes, masses and structure which would be very difficult to obtain otherwise.

We come now to those stars whose light variation is an intrinsic property of the star, a reflection of its fundamental constitution, and not just the result of external and fortuitous circumstances — one does not regard the

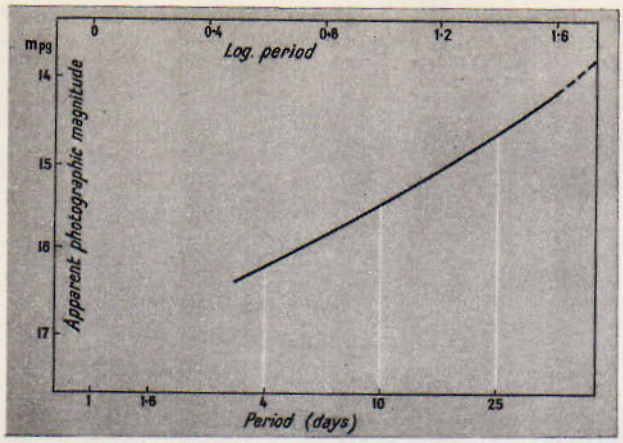


Fig. 580 — PERIOD/APPARENT BRIGHTNESS CURVE OF THE CEPHEIDS OBSERVED IN THE LESSER MAGELLANIC CLOUD BY MISS LEAVITT IN 1913

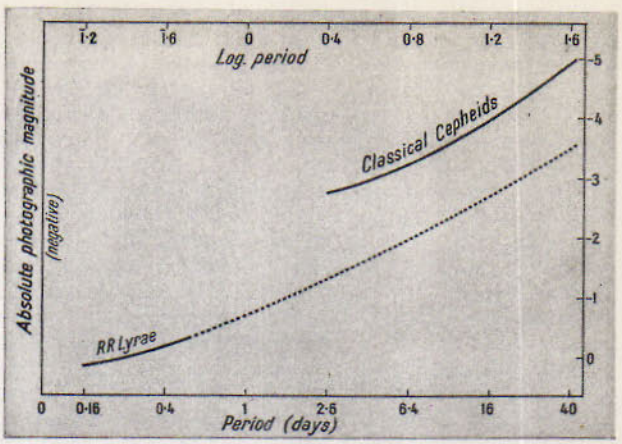


Fig. 581 — GENERAL CURVE FOR THE CEPHEIDS, AS ACCEPTED AT THE PRESENT TIME

Sun as a variable, for example, just because it is periodically eclipsed by the Moon. It may indeed happen that one of the components of an eclipsing binary is intrinsically variable (as in the case of VV Cephei), when the light curve will show superimposed fluctuations which make its analysis rather difficult. The events related in the next section are especially interesting, in that they form a remarkable example of the way in which recent discoveries have modified our former conceptions of the scale of the Universe, and of the way in which a few new facts can often provide the foundation for a hypothesis which supersedes its predecessors by eliminating inconsistencies left unexplained by them.

Cepheids and the Period-Luminosity Relationship. — δ Cephei is the brightest of this important class of variables, as well as being the first that was discovered (by Goodricke, 1784); hence their name.

Cepheids, of which more than 1,000 are known, are characterised by a light variation whose amplitude is about 1 magnitude, and which is performed with a clockwork regularity. The visual magnitude of δ Cephei, for example, varies from 3.6 to 4.4; its period of 5.37d is composed of 2d minimum-to-maximum and 3.4d maximum-to-minimum. Thus a rapid increase of brightness is followed by a more gradual decrease, giving the light curve (Fig. 579) the very characteristic Cepheid appearance.

However, this is not always so; in particular, a secondary maximum is frequently encountered, showing itself as a more or less well-defined bump on the light curve; this may occur on the descending or the ascending part of the curve, or even coincide with the primary maximum, which then appears abnormally sharp.

S. Gaposchkin, of Harvard College Observatory, has shown that as Cepheids of longer and longer period are considered, these irregularities are displaced steadily along the curves, and that at the same time the asymmetry of the curves increases. This, however, does not apply to the class of Cepheids as a whole, whose widely differing periods range from 1.5h (CY Aquarii) up to 2—3 months, but only within each sub-class into which they have been divided on the basis of their period distribution. Thus it has been discovered that Cepheids with periods of 5h, 1.5d, 9d, 27d and 60d are extremely rare, while those

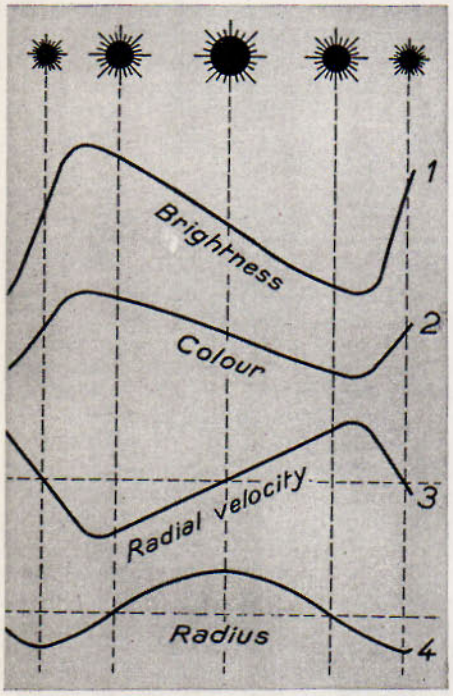


Fig. 582 — EXPLANATION OF THE LIGHT VARIATIONS OF THE CEPHEIDS. Above: variations in the diameter of the star. 1, variation in brightness; 2, in colour; 3, in radial velocity; 4, in radius.

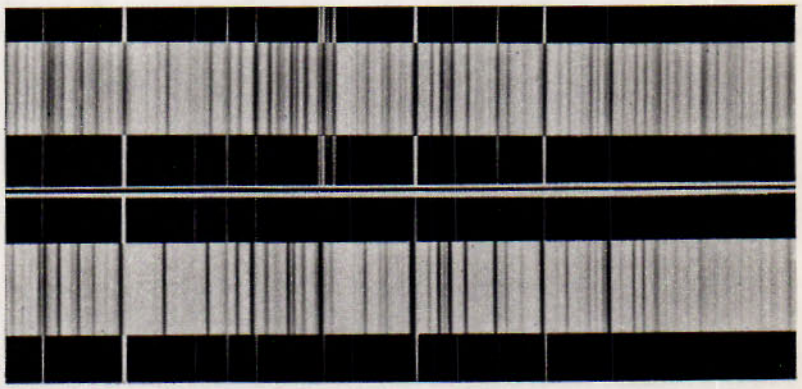


Fig. 583 — SHIFTS OF THE SPECTRAL LINES OF β CEPHEI, INDICATING A VARIABLE RADIAL VELOCITY AND THUS REVEALING THE STAR'S PULSATION. Above: 6 October 1907, radial velocity +2.4 m.p.s.; below: 7 October 1907, radial velocity −22.0 m.p.s.

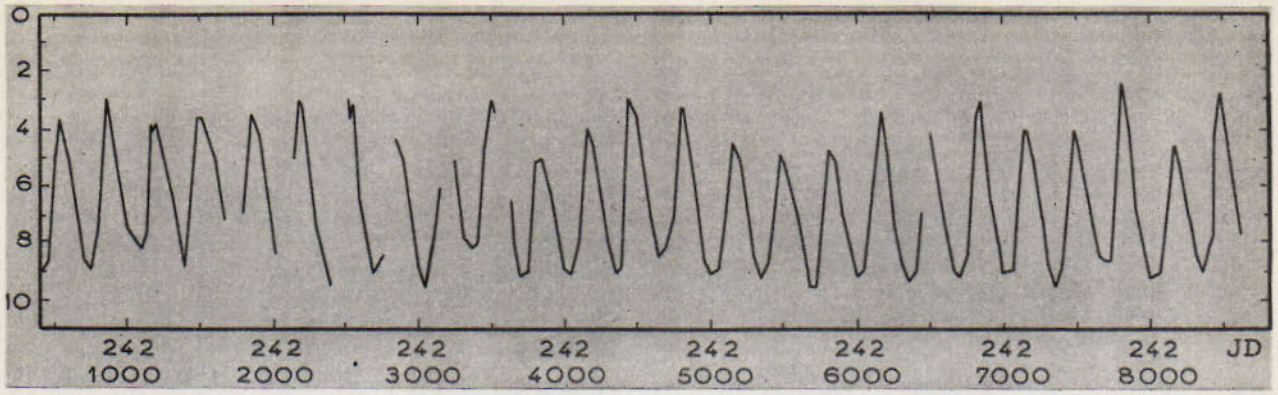


Fig. 584 — LIGHT CURVE OF A LONG PERIOD VARIABLE: MIRA CETI FROM 1915 TO 1936 (DIAGRAMMATIC). The abscissae are Julian Days, the usual time units in variable star work.

with intermediate periods — of the order of 3h, 12h, 5d, 13d and 45d — are numerous.

FEATURES OF SOME BRIGHT CEPHEIDS
(after C. and S. Gaposhkin)

STAR	MAGNITUDE*		PERIOD (DAYS)	SPECTROSCOPIC VARIATIONS
	MAX.	MIN.		
CY Aquarii	10.0	10.7	0.06	B8 — A3
RR Lyrae	7.2	8.0	0.57	B9 — F2
RT Aurigae	*5.4*	*6.6*	3.73	F1 — G5
α Ursae Minoris ...	*2.6*	*2.8*	3.97	F7 — F9
δ Cephei	*3.7*	*4.4*	5.37	F4 — G6
η Aquilae	3.7	4.4	7.18	F2 — G9
S Sagittae	*5.8*	*6.8*	8.38	F8 — G7
α Pavonis	4.2	5.1	9.11	F5 — G5
β Doradus	*4.2*	*5.7*	9.84	F2 — F9
ζ Geminorum	3.7	4.1	10.15	G1 var.
L Carinae	3.6	4.8	35.52	F8 — K0

* Photographic magnitude in italics

In practice the Cepheids are divided into two main classes: (a) short period variables of the RR Lyrae type (periods less than 1.5d), often known as cluster variables, from their numerous occurrence in globular clusters (Chapter 16); (b) the classical (1.5d to 27d) and long-period (greater than 27d) Cepheids.

It was at first presumed that these classes, however, did share a family relationship, and that the difference between the cluster variables and the classical Cepheids was nothing more than one of degree. Apparent support for this assumption was given by the discovery in 1913, by Miss Leavitt of Harvard College Observatory, of the so-called period-luminosity relationship.

It will be worth surveying the history of this in detail, for, although subsequent discoveries have to some extent invalidated Miss Leavitt's original findings, the technique for measuring great distances which this relationship gives us is still a powerful tool for astronomers.

At the time when Miss Leavitt put forward her suggestion, numerous Cepheids had already been discovered in the very distant star clouds, known as the Magellanic Clouds, which are situated in the southern hemisphere. When Miss Leavitt plotted the periods of these Cepheids against their mean apparent magnitudes, she found that the two quantities were related — the longer the period, the greater the mean apparent brightness of the star. But since these star clouds are of very small extent compared with the immense distance that separates them from us (of the order of 100,000 lightyears), all their constituent stars may for practical purposes be regarded as equidistant from the Earth. Hence the distance factor plays no part in their relative apparent brightnesses, which will thus differ from their intrinsic luminosities by a constant factor.

This relationship, thus established, was a purely empirical one. The possibility it offered was that of deriving the stars' distance in accordance with the simple formula which relates the apparent magnitude m and the absolute magnitude M* of a star:

$$m - M = 5 - 5 \log D$$

where D is the star's distance. If D (as in this case) is effectively constant throughout a cluster or cloud of stars, M will obviously differ from m by a constant interval for every member of the cluster.

So Miss Leavitt's discovery can be generalised as follows: if Cepheids are observed in any other star group, such as the Greater Magellanic Cloud, the nearest extra-galactic nebulae, etc., and for each such group a graph is set up whose abscissae are log P (period) and ordinates are m (apparent magnitude), then a series of smooth curves will be obtained all of which are of the same general shape; they will therefore be superposable merely by shifting them up or down the m axis.

It thus seemed to be firmly established that the observed Cepheids exhibit a definite period-luminosity relationship; these stars were all classical Cepheids with periods of from 3 to 45 days.

The various curves obtained from each cluster or cloud of stars (which differed from one another in the origin of their m scales, could be superimposed or placed end to end. But the fundamental problem then arose of how to calibrate the y-axis, allowing values of m to be converted to values of M, from which the distance D could be derived. Fig. 580 shows the curve representing Miss Leavitt's original observations, with the y-axis graduated in terms of m.

The problem of assigning M-values to the magnitude axis of the common graph combining all the classical Cepheids was a particularly difficult one, for the reason that no classical Cepheid is near enough to the Earth for its distance to be determined by either the trigonometrical or the spectroscopic method. Even δ Cephei, the brightest of the class, is more distant than 300 lightyears. It was therefore necessary to rely mainly on hypothetical parallaxes, the principle of which was explained in the last chapter, as established by the study of proper motions and radial velocities.

* The absolute magnitude of a star is the apparent magnitude that it would have if observed from a distance of 10 parsecs, or approximately 32.6 light-years. (See p. 336.)

Attention was then redirected to the Cepheids of very short period (the RR Lyrae variables) which are to be observed in large numbers both in the Galaxy and in the globular clusters.* These stars are characterised by relatively very large proper motions, which have been thoroughly investigated, and the mean distances of certain groups of RR Lyrae variables can be determined by the statistical methods which exploit the Sun's effectively uniform rectilinear motion towards the Apex.

Failing other and more precise methods, therefore, the various available Cepheid curves were either superimposed or adjusted end to end, and the results obtained from the lower end of the curve (corresponding to a period shorter than 0.8 days) were used to calibrate the entire curve.

Now all the measurements involving RR Lyrae Cepheid were in good mutual agreement, and indicated that the mean absolute magnitude of these stars is very close to zero. From this starting point, the distances of the classical Cepheids belonging to the Galaxy, the Magellanic Clouds, and also some of the nearest extragalactic nebulae, were deduced. These in turn allowed the distances of the more remote galaxies to be calculated.

The distance of the Andromeda nebula (M 31) was derived in this way as 300,000 parsecs, corresponding to a distance modulus ($m - M$) of 22.4:

$$m - M = 5 - 5 \log 300{,}000 - 22.4$$

But when the great Mount Palomar telescope came into operation, a disturbing fact obtruded itself on the notice of astronomers. The 'reach' of this instrument, expressed in apparent photographic magnitudes, is 23, and hence plates of the Andromeda nebula should be able to show RR Lyrae variables of apparent magnitude 22.4. Yet there was no sign of them. The inescapable conclusion was that M 31 must be significantly more distant than indicated by the Cepheid scale, as calibrated using RR Lyrae variables in our own Galaxy.

Walter Baade, of Mount Palomar Observatory, then tried employing another category of variable — the semi-variable whose type-star is Mira Ceti. These are very easily recognised from their spectra, which contain bright emission lines in addition to absorption lines. Their periods always exceed 90 days, and in 75% of cases lie between 200 and 400 days. The period of Mira itself ranges from 320 to 370 days, and its magnitude from 2 to 9.

Repeated and concordant measurements have assigned to these stars a tolerably general mean absolute magnitude of — 1.5. They have been identified on photographs of the Andromeda nebula, where their photographic magnitude is 22.4. Baade concluded from this that the distance modulus derived from the general curve of the Cepheids — which up to that time had been universally accepted — ought to be increased by 1.5 It follows from

$$5 \log D - 5 = k - 1.5$$

that log D should be reduced by 0.3 (the logarithm of 2). In other words the hitherto accepted distance of every extragalatic object would have to be doubled.

The accepted distances within the Galaxy itself would not, however, need revision except in the case of the individual classical Cepheids themselves, for these had not been used to determine the distances of any intragalactic star groups, clusters, etc.

The dimensions of the extragalactic nebulae must likewise be doubled, while their mean spatial density will be divided by 8. The constant of recession (Hubble's constant — see p. 422—426) will have to be reduced in the same proportion. This development, which was announced in September 1952, thus produced a far from negligible modification of our ideas concerning the Universe.

Since that date some independent confirmation has been obtained, notably by the discovery in the Magellanic Clouds of 26 RR Lyrae Cepheids — stars which it had never previously been possible to detect in the Clouds. And the tendency at the present time is to assume that the factor by which the 'old' extragalactic distances should be increased is in fact nearer to 3 than to 2 (General Assembly of the International Astronomical Union, Dublin, September 1955).

It should furthermore be noted that several considerations which were not new in 1952 have already led some astronomers to suspect that the absolute magnitudes of the Cepheids had been materially underestimated. Among these, the following may be mentioned:

(1) The study of the blue giants, novae and globular clusters that had been detected in M 31 and M 81;

(2) The value of the constant of recession, in theories of an expanding Universe (pp. 446—447), provides a span of past time since the onset of the expansion, which is insufficient to accommodate the known age of the Earth, and likewise the evolution of the stellar Universe;

(3) The previously adopted dimensions of the extragalactic nebulae implied that our own Galaxy is considerably larger than even the largest among them.

We now know that the two groups referred to earlier in this section — the 'cluster variables' and the classical Cepheids — do in fact belong to entirely distinct stellar populations.* The establishment of this fact is one of the most far-reaching in contemporary astronomy, and we shall discuss it in greater detail later.

Fig 581 shows the revised, discontinuous curve adopted today for Cepheids of all periods.

The fundamental cause of the Cepheid-type variation is still not perfectly understood. The discovery of apparent variations of radial velocity, synchronous with the magnitude variations, at one time led to the belief that the Cepheids are a particular sort of spectroscopic binary; but the comparison of the orbits derived on this assumption with the sizes of the stars concerned, quickly showed this hypothesis to be untenable — the radius of the former being considerably smaller than that of the latter, the stars would have to revolve inside one another. It is now generally agreed that (as originally proposed by the American astronomer Shapley, and the English astrophysicist Eddington) these variations of radial velocity are produced by the rhythmic pulsations of a single star, whose surface therefore actually advances towards and recedes from the terrestrial observer. Cepheids appear, then, to be pulsating stars, alternately inflating and deflating. Thorough analysis of the light and radial velocity curves, together with the study of the synchronous colour changes, has shown that the star attains its greatest brightness a short time before attaining its greatest size, for as it expands it cools, with consequent reduction of its surface brightness. The combination of these two factors determines the character of the observed light curve.

The variation of a Cepheid's diameter required to account for its observed magnitude range is, however, quite small — of the order of 10% of the whole diameter. It is not even certain that the entire mass of the star

* The distance of the nearest globular clusters is 7,200 parsecs, at which distance the traditional methods of distance determination are virtually useless.

* The gap occasioned in the series by the fact that no known Cepheid has a period between 0.8 and 3 days is an indication of this.

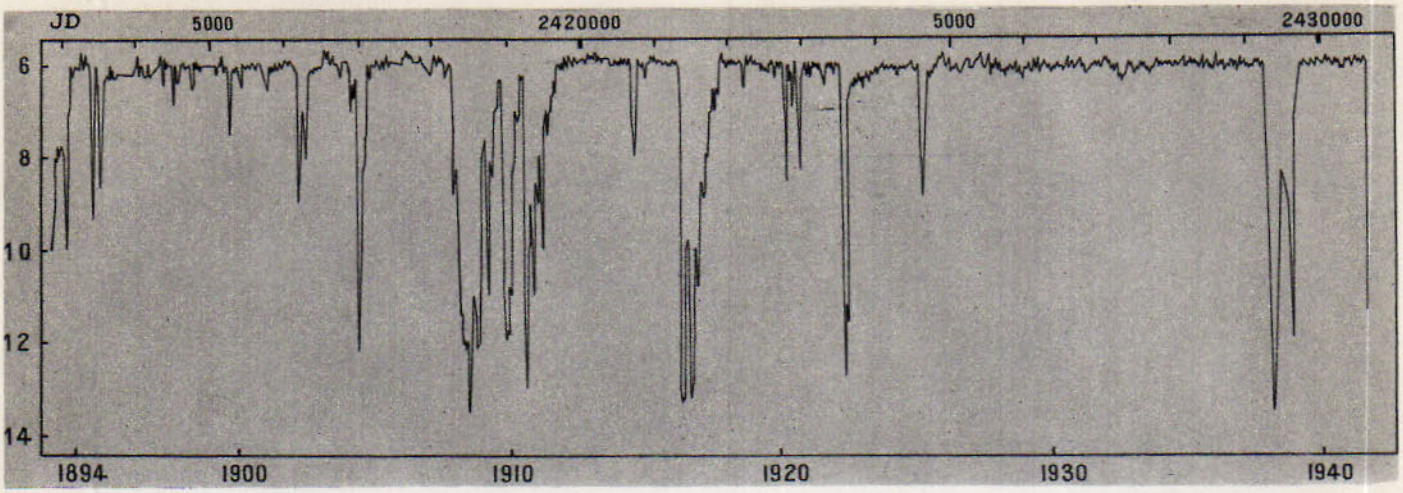


Fig. 585 — LIGHT CURVE OF THE IRREGULAR VARIABLE R CORONAE, FROM 1894 TO 1942. Above, date in Julian Days; below date in years. The 1945 fading continued to about magnitude 9 early in 1946. In 1948 there were signs of a new minimum approaching, and by early 1949 the star had experienced another major 'crisis', sinking to about magnitude 14. It remained at this level for the greater part of the year. Since then it has maintained more or less its normal level, with secondary minima at about magnitude 10 and 7 in 1953 and 1956 respectively.

participates in the pulsation, the modern view being that most probably it is the low-density surface layers that are chiefly involved. However, work still remains to be done on the detailed development of the theory of stellar pulsation.

A class of star that is often considered as being associated with the Cepheids is that whose type-star is β Canis Majoris. The amplitude of their light variation is very small (generally less than 0.1 magnitude) and can only be measured by sensitive photoelectric techniques. For this reason they have so far only been discovered among the brighter stars, but their relative numbers within this category show that their actual distribution in space must be very widespread. Their periods are also extremely short — normally between 3 and 8 hours — but the oddest thing of all about them is that the variation, after continuing regularly for a number of years with a perfectly well-defined periodicity, will die out completely for several years, during which time their brightness remains constant; then one day the tiny fluctuations will recommence, again following a well-defined period. The variation of γ Boötis is a typical example of this kind of behaviour. Furthermore, the radial velocities of these stars, which vary in step with the brightness, often show irregularities of form and amplitude when plotted as a curve. These variations, both of radial velocity and brightness, are attributed to very faint but more or less stable pulsations of the star's outer layers. Owing to their many points of resemblance with the Cepheids they are sometimes considered as 'dwarf Cepheids'.

Long-period variables. — We now come to an important class of variables whose periods are an extension of those of the Cepheids, ranging from 2 or 3 months to about 2 years, the most favoured periods being about 7 and 11 ½ months. Long-period variables are characterised by their colour, which is always reddish, indicative of relatively low temperatures (Cepheids are white or yellow), and by the large amplitudes of their brightness variations: these are usually in the neighbourhood of 3 to 4 magnitudes, though in certain cases they may be as small as 1 magnitude or as large as 10. This latter amplitude — exhibited by χ Cygni, whose magnitude range is 4 to 14, and whose period is 11 months — corresponds to a 10,000-fold variation of luminosity.

The type-star of the long-period variables is the well-known *o* Ceti (Mira); though discovered in 1596, the periodic nature of its variation was not realised till 1638. At maximum (magnitude 3) it is clearly visible to the naked eye, while at minimum (magnitude 9) it can only be seen with the aid of a small telescope; its mean period is 11 months, of which the rise from minimum to maximum occupies 4—5, and the fading from maximum to minimum 6—7.

These oscillations, moreover, are very far from sharing the clockwork-like regularity of the Cepheids: amplitude, period and shape of the light curve are all subject to continuous modifications, as is shown by Fig. 584. Often Mira is no brighter than magnitude 4 at maximum, while at other times it may attain magnitude 2; once, indeed, in 1779, it reached the first magnitude and rivalled Aldebaran (α Tauri). On the other hand, it has been as faint as magnitude 5 at maximum — in 1924 and 1935, for example, when it was barely visible to the naked eye. The minima are also variable, though to a smaller degree than the maxima, their outside limits being magnitudes 8 and 10.

FEATURES OF SOME LONG-PERIOD VARIABLES
(after C. and S. Gaposhkin)

STAR	MEAN MAGNITUDE		MEAN PERIOD (DAYS)	SPECTROSCOPIC VARIATIONS
	MAX.	MIN.		
L Puppis	4.0	6.0	140	M5*e* — M6*e*
S Carinae	5.3	9.3	149	K7*e* — M4
W Orionis	5.9	13.6	200	N5
R Trianguli	6.0	11.5	266	M4*e* — M8*e*
RR Scorpii	6.0	11.5	279	M5*e* — M8*e*
R Carnae	4.4	9.4	308	M4*e* — M7
R Leonis	5.9	10.1	313	M6*e* — M9*e*
R Aquilae	6.0	11.4	329	M5*e* — M8*e*
o Ceti	3.4	9.2	332	M5*e* — M8*e*
U Orionis	6.0	11.8	372	M6*e* — M8*e*
T Cephei	6.1	10.1	387	M5*e* — M7*e*
R Aquarii	6.4	10.3	387	M6*e* + P
R Hydrae	4.2	9.5	405	M6*e* — M9*e*
χ Cygni	5.1	13.3	407	M6*e* — M8*ep*
R Centauri	6.0	11.0	564	M4*e* — M5*e*

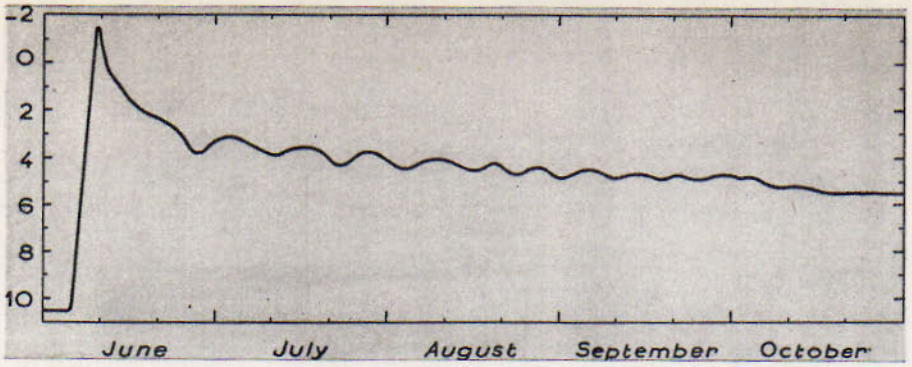


Fig. 586 — LIGHT CURVE OF NOVA AQUILAE 1918

Similarly, the interval between successive maxima or minima is far from maintaining a constant value: it may be as short as 320d or as long as 370d, and despite the quantity of statistical work that has been devoted to discovering some regularity in these variations, they appear to be subject to no ascertainable law. For this reason, the times of maximum or minimum of the long period variables cannot be predicted with an accuracy exceeding, on the average, 10 or 15 days, and they have to be kept under constant observation. For over half a century this has been the particular preserve of various associations of amateur astronomers, one of the more important of which is the American Association of Variable Star Observers, whose members have accumulated more than a million observations (mostly of long-period variables) during the past forty years. The Variable Star Section of the British Astronomical Association also deserves mention: since its foundation in 1891, its members have recorded some 400,000 individual observations, of which over 300,000 refer to long-period variables.

The behaviour of many long-period variables is even more irregular than that of Mira. Their light curves are frequently distorted by a more or less obtrusive hump, whose position does not even remain constant from one period to the next (e. g. χ Cygni); it may at times be so marked as to produce what is in effect a secondary minimum, giving the curve a very characteristic shape (e. g. R Centauri).

The main forms taken by long-period variable light curves are now well known, but unfortunately this cannot be said of the cause of their variation. Spectroscopic investigation has shown that the variation of the visual brightness is for the most part due to large-scale fluctuations in the absorption exercised by titanium oxide vapour — a compound which occurs abundantly in the atmospheres of this type of red star, and whose absorption is restricted almost entirely to the visual wavelengths; these variations in turn result from temperature changes. We have seen that the lowering of the temperature favours the formation of molecules (including those of titanium oxide); it could be responsible for the development of immense clouds of absorbing material which would effectively obscure the surface of the star. It has been discovered that, outside the specific titanium oxide absorptions, the variation of the infra-red emission of these stars is much smaller than that in the visual region of the spectrum, which suggests that the temperature variations are relatively small. As to the origin of the temperature variations themselves, it is generally supposed that — as in the case of the Cepheids — they are caused by slight but more or less regular pulsations of the gaseous material of the star, though the theory of these pulsations is much less advanced than that of the Cepheids.

Semi-regular and irregular variables. — The behaviour of the long period variables is, as we have seen, not perfectly regular; this characteristic becomes still more marked among the semi-regular variables, whose light curves nevertheless exhibit a recognisable periodicity of greater or smaller permanence. This class is also composed mainly of red stars; their brightness varies in a more or less regular manner, over a range (normally) of about a magnitude, and exhibits pseudo-periods of the order of 2 to 4 months.

Sub-classes can be differentiated, some fairly clearly, though their boundaries often tend to be vague. The chief of these are represented by the type-stars (*a*) RV Tauri: light curves often reminiscent of very erratic β

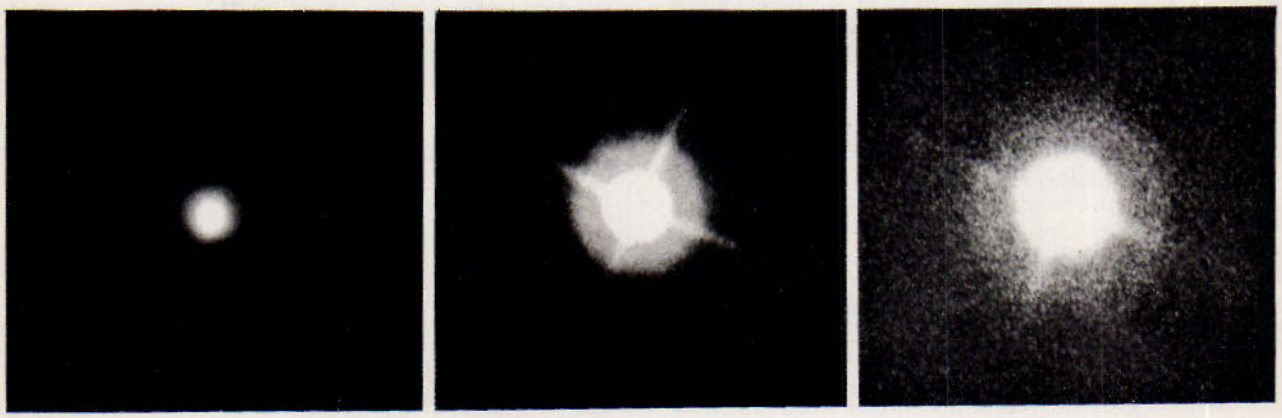


Fig. 587 — PROGRESSIVE EXPANSION OF A NEBULOUS SHELL AROUND NOVA AQUILAE 1918. Left to right: 20 July 1922; 3 September 1926; 14 August 1931. *(Photos, Mount Wilson Observatory, 100-inch reflector)*

Lyrae-type eclipsing binaries; (*b*) SX Centauri: very curious light curves, consisting of rapid, fairly regular, short-period oscillations (15 to 45 days) of small amplitude (1 to 2 magnitudes), on which is superimposed a longer periodicity (2 to 3 years) of larger amplitude (2 to 5 magnitudes); (c) μ Cephei: somewhat similar to the preceding, but even less regular and of smaller amplitudes; another example is the beautiful first magnitude red star α Orionis (Betelgeuse), with a principal period of about 6 years superimposed on innumerable minor fluctuations between the magnitude limits of about 0 and 1; it is thus the peer of Arcturus or Capella at maximum, but only about equal to Aldebaran or Pollux (β Geminorum) at minimum.

ELEMENTS OF SOME SEMI-REGULAR VARIABLES
(after C. and S. Gaposhkin)

STAR	MEAN MAGNITUDE		MEAN PSEUDO-PERIOD (DAYS)	MEAN SPECTRO-SCOPIC TYPE
	MAX.	MIN.		
T Centauri ...	6.1	7.7	91	Mo
U Monocerotis	5.6	7.3	92	G8
X Herculis ...	5.8	7.2	100—400	Mc
α Herculis ...	3.1	3.9	120—2,250?	M5
W Cygni	5.1	7.0	131	M4
T Ceti	5.2	6.0	159	M3
W Orionis....	5.9	7.7	200—2,000?	N5
UU Aurigae .	5.1	6.8	235—3,000	N3
RS Cancri....	5.3	6.8	239	M6
μ Cephei	4.0	4.8	750	M2
ϱ Cassiopeiae .	4.4	5 1	1,100	G5
α Orionis.....	0.1	1.2	2,070	M2

Fig. 588 — The gases expelled from Nova Persei 1901 had by 1917 formed an extremely asymmetrical nebulosity. *(Photo, Mount Wilson Observatory, 60-inch reflector; 5½ hours exposure)*

Finally, after all traces of periodicity have in theory been swamped in chaotic irregularity, we come to the class of irregular variables; even among these, however, we can sometimes detect traces of periodicity (several months) over fairly long periods.

Some of these are hardly variables at all — or at any rate only vary very infrequently, and after long periods

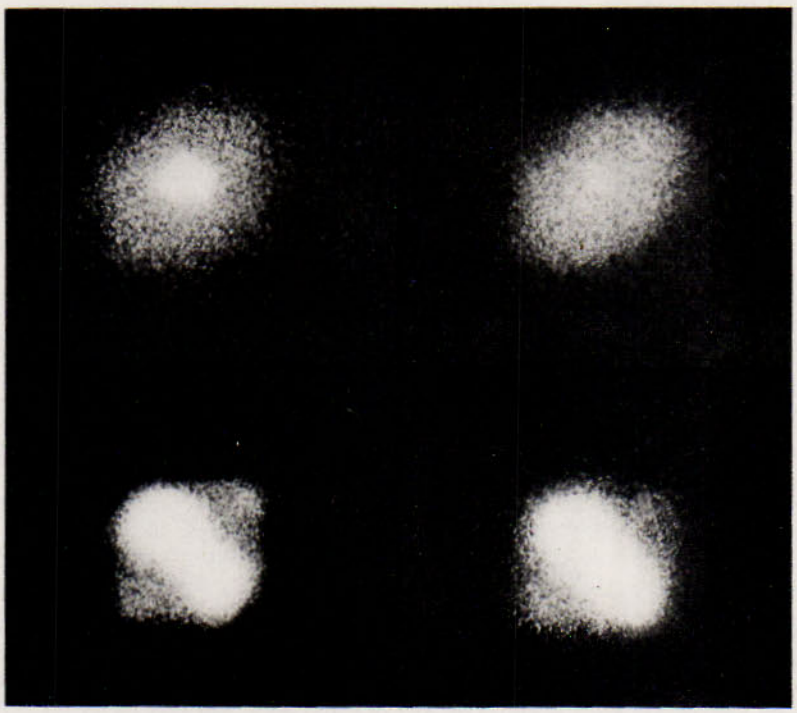


Fig. 589 — DIFFERENT DISTRIBUTIONS OF THE CHEMICAL ELEMENTS IN THE ENVELOPE SURROUNDING NOVA HERCULIS 1934, AS OBSERVED IN 1942. Top left, the star and its envelope in blue-violet light; right, the envelope alone in the light of the green N1 and N2 lines of doubly ionised oxygen; lower left, the envelope alone in the light of the red lines (λ 6584, 6548) of ionised nitrogen, July 1942; right, the same, August 1942. *(Photos, Mount Wilson Observatory)*

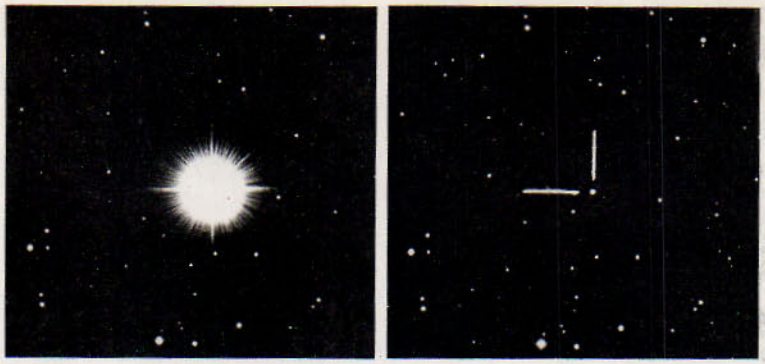


Fig. 590 — NOVA HERCULIS IN 1935. Left, near maximum, 10 March 1935; right, at minimum, 6 May 1935. *(Photo, Lick Observatory, 36-inch telescope)*

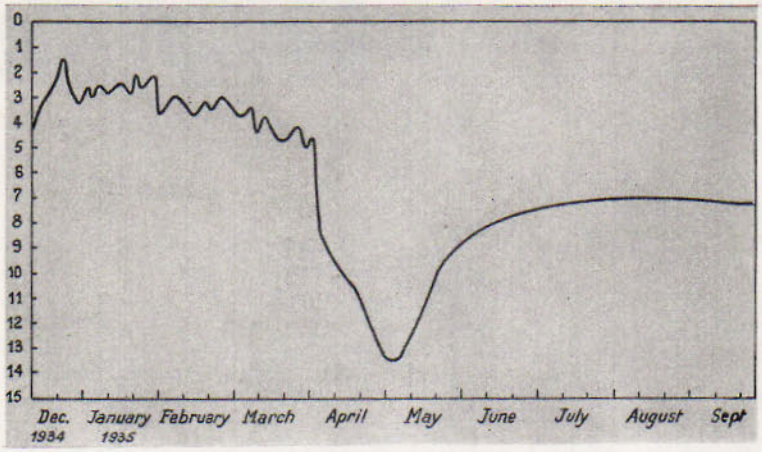


Fig. 591 — LIGHT CURVE OF NOVA HERCULIS 1934

SOME IRREGULAR VARIABLES (after C. and S. Gaposhkin)

STAR	MAGNITUDE		SPECTROSCOPIC TYPE
	MAX.	MIN.	
ϱ Persei	3.2	4.1	M*b*
R Lyrae	4.0	4.5	M5
α Ophiuchi	4.1	5.0	K0
g Herculis	4.4	5.6	M*bp*
U Hydrae	4.8	5.9	N2
UU Aurigae	5.1	6.8	M3
SX Pavonis	5.3	6.3	M*b*

of inactivity. Such is the case with α Cassiopeiae, a magnitude 2 star whose supposed variability seems in fact to be almost wholly attributable to errors of observation. Others, no doubt, are really semi-regular variables whose periodicity has not yet been disentangled, for these stars are difficult to observe — owing to their red colour, which is often very pronounced — and are still inadequately investigated.

The cause of their variations is not yet known, though it may be analogous to that already described in connection with the long-period variables; like the latter they frequently, if not invariably, exhibit roughly simultaneous variations of radial velocity and of the intensity of their molecular absorptions (titanium oxide, carbon, cyanogen, as the case may be) which are capable of modifying the opacity of their atmospheres to a very marked degree. But we are still far from a proper understanding of the processes involved.

Peculiar variables. — Under this head are gathered for convenience, a number of stars which are often classed as irregular or semi-regular variables; some at least, however, have analogies with the 'temporary stars', or novae, which we shall consider next.

The type-star of the first group of these peculiar variables is R Coronae — the brightest and most thoroughly investigated of a small class of stars, only a few score of which have been identified with certainty. R Coronae is normally of the sixth magnitude — just about the limit of naked-eye visibility — and may maintain a constant brightness for months or even years (1926–34, for instance) without significant variations. But from time to time it will fade, suddenly and without warning, and usually to about magnitude 12; it will then regain its original magnitude quite quickly (a matter of months), though its rise is marked by large fluctuations; here, at magnitude 6 once more, it will remain stable until the next 'crisis' occurs.

Its most recent fadings were in 1917–8, 1923, 1925, 1938–9, 1942–3 and 1945 (Fig. 585), but no one can predict the next.

One characteristic feature of these stars is the abundance of carbon and its compounds in their atmospheres, and it is possible that there is some connection between this and their very typical fluctuations. Small temperature changes may perhaps induce the condensation of atmospheric carbon, for the temperatures of these atmospheres are in the region of the condensation temperature of this element. In any case it is doubtful whether their transitory obscurations can be explained by means of the mechanism attributed to the long-period variables.

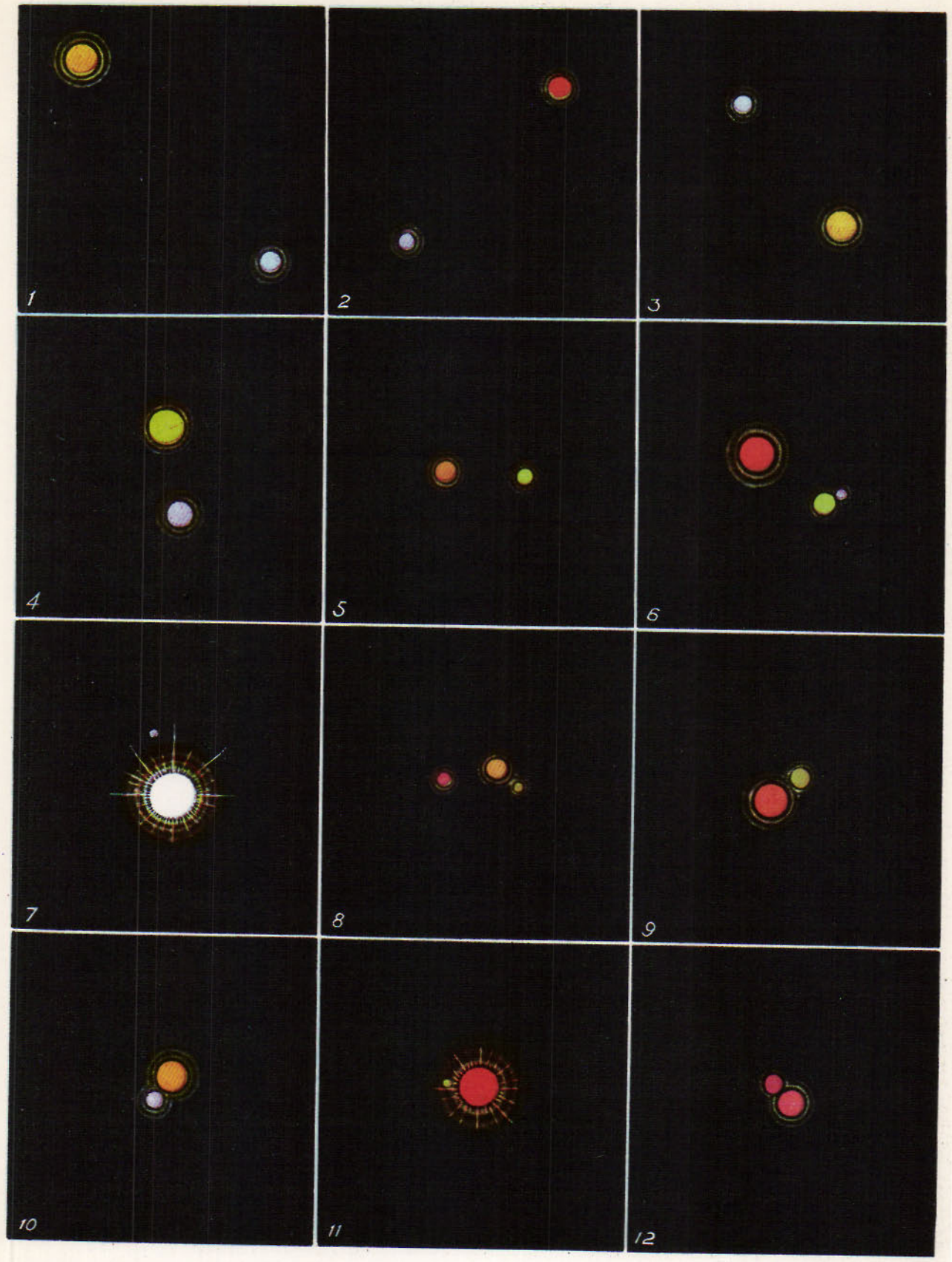


Plate XI. COLOURED MULTIPLE STARS

1. β Cygni, 2. η Persei. 3. α Canum Venaticorum. 4. β Scorpii. 5. γ Delphini. 6. γ Andromedae. 7. β Orionis. 8. ι Cassiopeiae. 9. α Herculis. 10. ε Boötis. 11. α Scorpii. 12. ξ Boötis.

N. B. For technical reasons the apparent diameters of the discs and diffraction rings of the brightest stars, and also their colours, are exaggerated considerably.

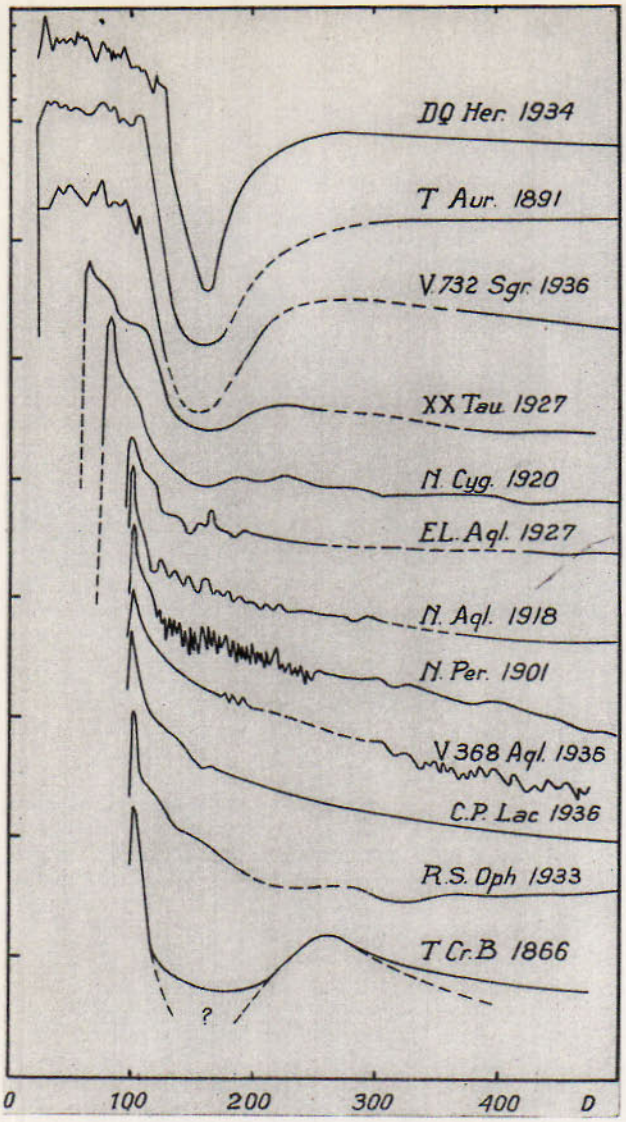


Fig. 592 — SUGGESTED CLASSIFICATION OF THE LIGHT CURVES OF NOVAE (after L. Campbell, Harvard Observatory).

It has been suggested — and is by no means impossible — that these stars appear variable to us because of their accidental occultations by obscuring clouds in space, behind or through which their proper motions are carrying them.

This suggestion naturally leads us on to consider those stars which are visibly associated with nebulosity, and whose variations appear to be directly connected with this diffuse interstellar medium. Many such stars — for which C. and S. Gaposchkin of Harvard College Observatory have proposed a special sub-class of 'nebular variables' — have no more than this single feature in common, for they tend to exhibit the most diverse behaviour patterns. Stars that truly qualify for inclusion in this category are the numerous irregular variables associated with the great Orion nebula (see Chapter 16), those that are immersed in the bright and dark nebulous regions of Carina, Monoceros and Corona Australis, and, finally, the principal stars of the Pleiades, which from time to time undergo slight but characteristic variations of brightness. They are all bluish, high-temperature stars whose intense radiation lights up the clouds of dust or gas in which they are immersed, and their variability — or at

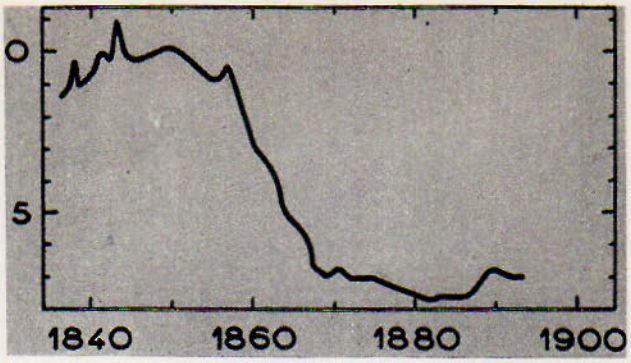


Fig. 593 — THE LIGHT CURVE OF η CARINAE DURING THE NINETEENTH CENTURY

any rate a great part of it — may be attributed to the varying opacity of the cosmic cloud through which we happen to be viewing them.

On the other hand there are also red, low-temperature stars which are associated with nebulosity (R Aquarii, for example), but it is suspected that in reality these are often doubles (as Z Andromedae) with a high-temperature *comes* — an unusual stellar association which is also encountered among the long-period variables (Mira, for instance) and 'temporary stars' (T Coronae Borealis).

Another class of nebular variable includes some interesting, very high-temperature stars which appear to produce their nebular envelopes by sloughing off their outer layers. These ejected layers are not normally recorded on direct photographs of the night sky since they tend to be very close to the primary, about which they form a kind of shell; they are, however, detectable by certain variable characteristics of the star's spectrum (hydrogen emission lines). Furthermore, the total brightness of a good number of these stars is not sensibly variable — or, at least, appears to remain constant for long periods; but sooner or later the spectroscopic variations will set in, and then the observed brightness of the star will follow suit.

The most remarkable example of this type of star is the second-magnitude γ Cassiopeiae. Until 1936 it had shone steadily, but early in that year it suddenly and unexpectedly started to brighten; this continued until

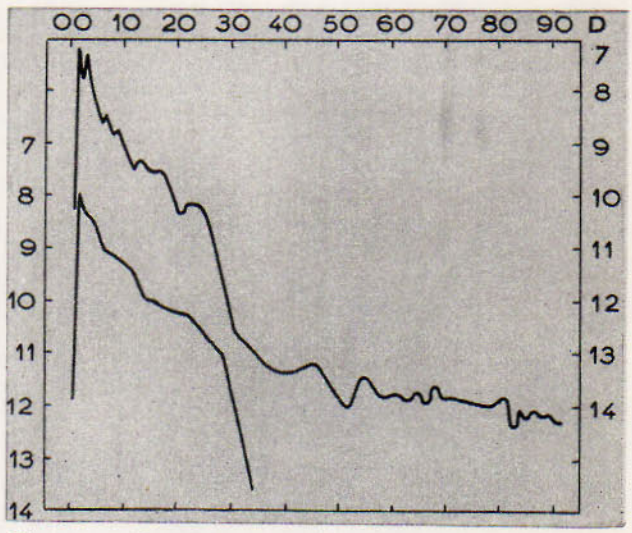


Fig. 594 — SIMILARITY BETWEEN THE LIGHT CURVES OF THE RECURRENT NOVA T SAGITTAE IN 1913 (UPPER) AND 1946 (LOWER)

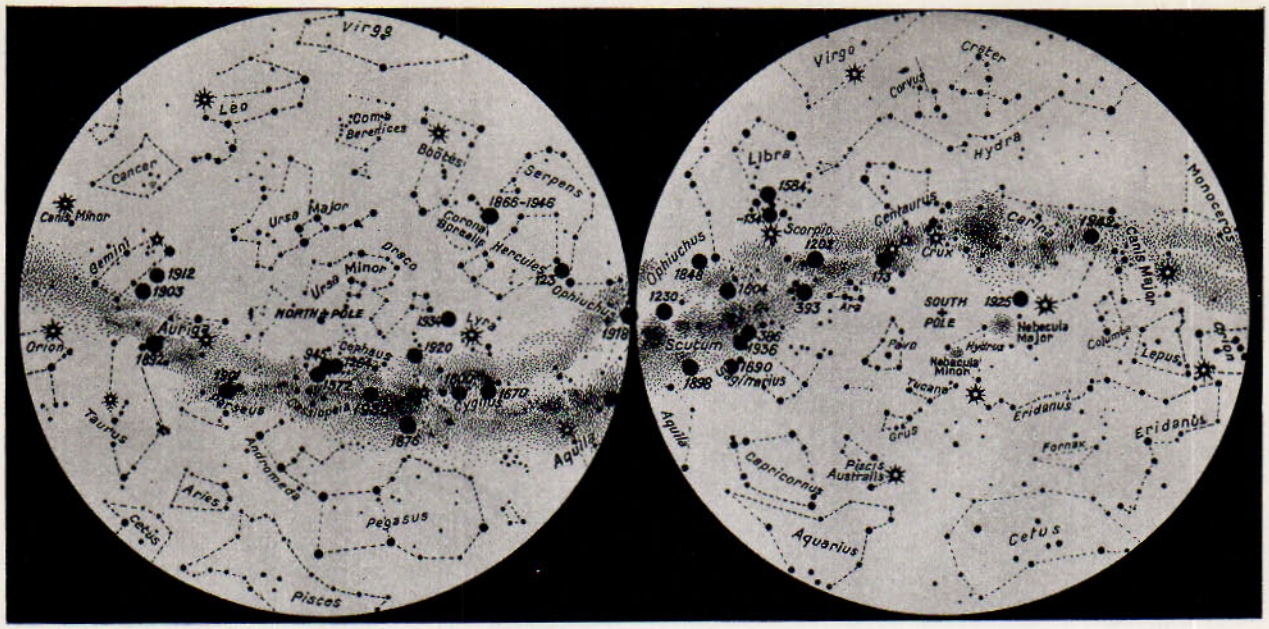


Fig. 595 — DISTRIBUTION ON THE STAR SPHERE OF THE BRIGHTEST NOVAE OBSERVED DURING THE LAST 2,000 YEARS

May 1937, by which time it had reached magnitude 1.4 (i.e. twice its original brightness of magnitude 2.2) and had deepened in colour to an abnormal yellow tint. During the later months of 1937 it faded rapidly, with fluctuations of decreasing amplitude, and by 1940 it had sunk to a sensibly lower level than that from which it had started (about magnitude 2.8); here, more or less constantly, it has remained since 1941. Very accurate measures have, however, shown that faint and quasi-periodic variations still persist, and there is no reason to suppose that its more spectacular performance may not one day be repeated.

A large number of spectroscopic observations were made during its outburst, and these indicate that for some unknown reason the star expelled its outer atmospheric and photospheric layers; their expansion provided both the increase in brightness and the change of colour (since the gases cooled as they expanded); ultimately this envelope became entirely detached from the star, whose new photospheric surface then became visible.

Such, in outline, is the picture that has been built up of the cataclysms that are liable to overtake these high-temperature stars with emission spectra. Among them, also, is included the class of P Cygni variables, whose increases of magnitude are in some cases sufficiently spectacular for them to be considered as true novae. This was the case with the type-star, now a faint and slightly variable object of the fifth magnitude, which at least twice during the last 3½ centuries (1600 and 1657) has brightened to the third magnitude.

Novae, or temporary stars. — These are stars which suddenly break away from the anonymity that surrounds the myriads of faint stars, and increase rapidly in brightness until they often outshine the brightest stars in the sky. Once maximum has been reached they fade again, more slowly and suffering many minor fluctuations of brightness, until, perhaps 20 or 30 years after the outburst, they have regained their original brightness; thereafter they continue to vary irregularly over a small range.

The name stella nova ('new star') was given to this type of object in pre-telescopic days, when it did indeed appear that they loomed up out of nothingness, subsequently to disappear again without a trace. A perhaps more suitable name would be 'temporary star' — the implication being understood that it is the star's outburst, rather than the star itself, that is temporary.

One of the finest examples of recent times was Nova Aquilae 1918. This magnitude 11 star — which, without being in any way noticeable, had been recorded on many earlier photographs — on 7 June 1918 suddenly experienced an explosive outburst that shot it to magnitude 6 the same evening and to magnitude 1 by the following evening. Still brightening, it attained magnitude −1.5

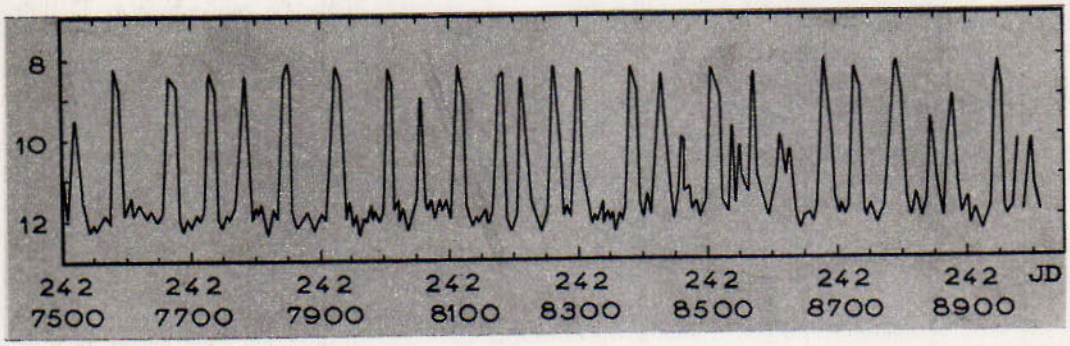


Fig. 596 — LIGHT CURVE OF SS CYGNI FROM 1930 TO 1935 (based on observations of the American Association of Variable Star Observers)

for the space of a few hours during the night of 9/10 June, when it outshone every other star visible at the time. Its brightness had increased 70,000 times in three days; but its splendour was short-lived, for by the following night it had already faded to magnitude 0, and eight days later to the second; by the end of the month it was between the third and fourth magnitude. Its rate of fading then slowed down considerably, while its fluctuations of brightness became progressively smaller. By September 1918 it had sunk to magnitude 5, after which its fading continued slowly and steadily, carrying it below the naked-eye threshold by the end of the year, and through the seventh magnitude during 1919 to magnitude 8 at the beginning of 1920, 9 at the beginning of 1921, and 10 during 1922; finally it reached its original magnitude (10.5) in 1924–5, and there it has remained, though varying irregularly through a range not exceeding 0.5 magnitude.

Spectroscopic observations of Nova Aquilae revealed high radial velocities — more than 600 m.p.s. at maximum — which lend weight to the idea of an explosive expansion of the photosphere and atmosphere. Furthermore, the expelled gases soon became observable telescopically as a tiny nebulous shell whose diameter continued to increase as the brightness decreased. It was first observed visually by Barnard with the great 40-inch refractor at Yerkes Observatory (near Chicago) towards the end of 1918, when its angular diameter was only 2″. It had increased to 4″ by 1920, when it became too faint for visual observation, although photography recorded its continued increase to 16″ diameter by 1926 and 44″ by 1940 — a steady expansion of 2″ per year (Fig. 587). This small nebula must now be well over 1′ in diameter, but it can only be detected with difficulty owing to its exceeding faintness. However, in 1941 Baade succeeded in photographing it in red light at Mount Wilson, and subsequently employed the same technique to discover nebulous envelopes surrounding the ex-novae Aurigae 1892, Cygni 1920 and Lacertae 1936. Similar nebulosity, markedly asymmetrical, is also to be seen around Nova Persei 1902 (Fig. 588).

Fig. 597 — SUPERNOVA WHICH APPEARED IN THE EDGE-ON SPIRAL NEBULA NG 5907 IN 1940. Left: Mount Wilson Observatory, 60-inch reflector, enlarged 1.5 times; right: Mount Palomar Observatory, 18-inch Schmidt, enlarged 12.8 times.

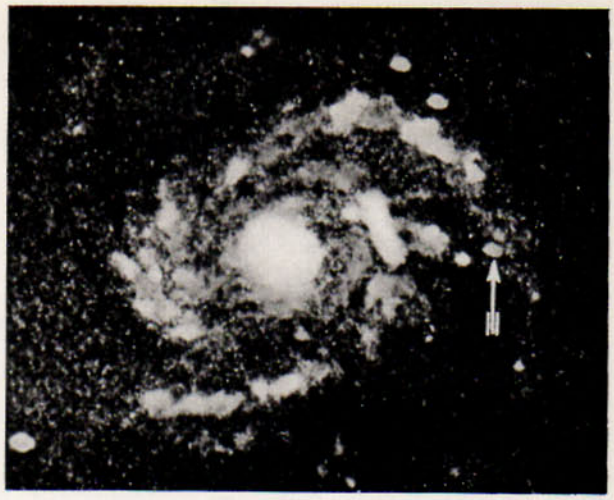


Fig. 598 — SUPERNOVA WHICH APPEARED IN THE SPIRAL NEBULA M 100 IN 1901. *(Photo, Lick Observatory; 36-inch telescope)*

If it can be assumed that, as is indicated by its spectrum, the radial expansion of the Nova Aquilae gaseous shell was accomplished at a velocity of 1,000 m.p.s., an approximate distance of 1,000 light years can immediately be derived. At such a distance the star would have an absolute magnitude of −9 at maximum and +3 at minimum. In fact, it is known that the gases are not expelled uniformly by a nova, but primarily in two favoured and diametrically opposed directions. By chance these happened to coincide with the line of sight in the case of Nova Aquilae 1918; in other cases, however, they are seen at right angles to our line of sight, when the expelled gases are sometimes observed as an elongated nebulosity expanding to form a more or less elliptical halo round the nova. At other times (or even simultaneously) two or

THE BRIGHTEST NOVAE (Max. mag. < 6) OBSERVED BETWEEN 1843 AND 1951

NO.	YEAR	NOVA	MAGNITUDE	
			MAX.	MIN.
1	1843	η Carinae	− 1.0	7.8
2	1848	Ophiuchi II.....	5.0	13.0
3	1866	T Coronae	2.0	10.6
4	1867	Q Cygni	3.0	15.4
5	1877	Comae (?)	4.5	> 10
6	1891	T Aurigae	4.2	15.0
7	1898	Sagittarii I	4.7	> 16
8	1901	RS Ophiuchi III ...	4.3	11.8
9	1901	Persei II	0.0	14.0
10	1910	Lacertae I	5.0	14.4
11	1912	Geminorum II	3.7	14.5
12	1918	Aquilae III	− 1.4	10.8
13	1918	Monocerotis I	5.4	15.1
14	1920	Cygni III.....	1.5	> 16
15	1925	RR Pictoris	1.0	12.7
16	1933	RS Ophiuchi III ...	4.3	11.8
17	1934	DQ Herculis	1.3	15
18	1936	CP Lacertae	2.2	15
19	1936	Sagittarii	4.5	> 15
20	1939	Monocerotis II	4.3	15.8
21	1942	Puppis	0.1	> 14.6
22	1946	T Coronae	1.7	> 10.6
23	1948	Serpentis	4	> 14
24	1950	Lacertae II	5	> 16

more small diffuse spots may develop alongside the nova, and, gradually separating, give the impression of double or multiple offspring of the explosion. Nova Pictoris 1925 presented such an appearance, which was even more marked in the case of Nova Herculis 1934. Later observations of the latter star have even established that the various chemical elements are not uniformly distributed throughout the gaseous envelope (Fig. 589).

The absolute magnitudes of novae at maximum are difficult to determine with accuracy since none has yet been detected which is sifficiently near to show a trinogometrical parallax of any size. But various indirect methods (described in the last chapter) have allowed their order of size to be determined well enough to show that at maximum they are more luminous than any normal star: the value of the mean absolute magnitude at maximum which is generally accepted today is roughly -6 or -7, which is equivalent to from 20 to 50 thousand times the luminosity of the Sun. In their normal state (minimum), on the other hand, their absolute magnitudes of $+4$ to $+5$ are comparable with that of the Sun.

Nevertheless, all novae are far from sharing identical characteristics or exhibiting a like behaviour, as can be seen from the table of the principal novae which have appeared during the last hundred years. Their amplitudes, though generally about 11 to 12 magnitudes, may be anything between 10 and 13. On the other hand their light curves have certain very well-marked traits, by means of which they may be subdivided into a number of fairly definite classes. The so-called 'rapid' novae are characterised by an extremely sudden rise to maximum — accomplished in two to three days or less — followed by an initially rapid fading which later slows down into a progressively less steep decline; this may be relatively smooth (N Cygni 1920, N Lacertae 1936, N Puppis 1942) or marked by a long series of fluctuations (N Persei 1901, N Geminorum 1912, N Aquilae 1918). Sometimes, too, a rapid intitial fading will be followed by a temporary recovery to a secondary maximum before the final decline supervenes (e.g. T Coronae Borealis 1866, 1946). This temporary recovery is especially characteristic of the second group, the 'slow' novae, whose fairly steady rise to maximum often occupies 10 to 15 days. Further, they often remain at or near maximum brightness for several months, fading very gradually and with many random fluctuations; then a more rapid phase of the decline sets in, and within a few days they are back at their original magnitude. However, they do not remain at this level, but brighten again over a period of several weeks to a secondary maximum, from which they return slowly and finally to their minimum.

Novae Aurigae 1891 and Herculis 1934, as well as those observed in Sagittarius in 1936 and Cygnus in 1942, exemplified very clearly this particular type of light curve. Other slow novae do not have the temporary relapse, but fade steadily, slowly, and with many irregular fluctuations from maximum to minimum (N Pictoris 1925, N Aquilae 1936 I).

Finally, there is a class of stars whose light curves are akin to those of novae in shape, but whose slow-motion performances may extend over some scores of years. One of the best studied examples is RT Serpentis: prior to 1909 it was unknown, being fainter than magnitude 16; during that year it suddenly began to appear on photographs with a magnitude of about 11.5; its slow rise continued till it reached a maximum of about magnitude 10.5 which it maintained for five years (1919–23); it then faded, with many brightness fluctuations, to about magnitude 13, where it has remained since 1930. Another example of this class appeared in Orion in 1939.

But the most remarkable one — by virtue both of the slowness and the amplitude of its variation — is η Carinae, which had been variously described as a 4th and as a 2nd magnitude star at different times during the 17th and 18th centuries, and then at the beginning of the 19th century underwent a spectacular brightening through magnitude 4 in 1810, 2 between 1820 and 1830, finally becoming almost the peer of Sirius in 1843. After this maximum it faded to magnitude 1 by about 1854; then from about 1857 the decline became more rapid; by 1860 it was of magnitude 3 and it became invisible to the naked eye in about 1867. Three years later it reached magnitude 7 and has remained at this level, though with interesting fluctuations, ever since (Fig. 593).

η Carianae appears to be involved in a large diffuse nebula, and may therefore perhaps be considered as a nebular variable. On the other hand its spectroscopic changes would seem to ally it with the P Cygni variables.

Recurrent novae, and the SS Cygni variables. — Novae are very much more numerous and more frequently occuring than the rarity of the most spectacular examples of the class might lead one to believe. Since the inception of a virtually continuous photographic patrol of the heavens (mainly by Harvard College Observatory) hardly a year has passed without one or more novae being detected on the plates. About 100 have in this way been recorded during the past half century, but a still greater number must escape notice, being lost among the vast multitudes of stellar points that are piled up on the negatives. Bearing this in mind, it has been estimated that at least 20 or 30 novae must occur annually.

From this it follows, either that all the stars in our Galaxy must have passed through a nova stage during the course of the last hundred thousand years, or else that (if only a certain type of star is capable of becoming a nova) the same stars must experience the nova outburst on more than one occasion. In the latter case, the same star would reappear as a nova from time to time, at longer or shorter intervals.

In recent years this has in fact been shown to happen, and an increasing number of so-called recurrent novae have been discovered and studied. Thus the nova T Coronae of 1866 reappeared in 1946, repeating almost exactly its behaviour of 80 years previously, while nova T Sagittae of 1913 likewise gave a repeat performance in 1946 (Fig. 594). RS Ophiuchi has also been observed at two maxima (though these were admittedly unequal), in 1901 and 1933, and U Scorpii has been observed at three — 1862, 1906 and 1936. Lastly, T Pyxidis has already exhibited four well-observed outbursts, in 1890, 1902, 1920 and 1945.

This phenomenon of recurrence demonstrates two things: that the possibility of experiencing the explosive outburst which constitutes a nova is limited to a very definite type of star; and that only the external layers of the star are expelled, leaving the star itself relatively unscathed and ready to repeat the performance. Moreover, as the American astronomer McLaughlin has pointed out, the amplitudes of recurrent novae are smaller than those of normal novae, which we have only seen explode once — i.e. whose successive maxima are separated by very much longer intervals.

This characteristic has led some astronomers to associate the recurrent novae with the small but remarkable class of SS Cygni or U Geminorum variables, of which only a few score are known. At irregular intervals of from roughly 20 to 150 days these stars undergo sudden increases of brightness, amounting on the average to 3 or 4 magnitudes, then rapidly fade again, usually in a matter of days, to their normal level; there they remain, usually without significant changes, till the next outburst. The Z Camelopardalis stars constitute a subdivision of

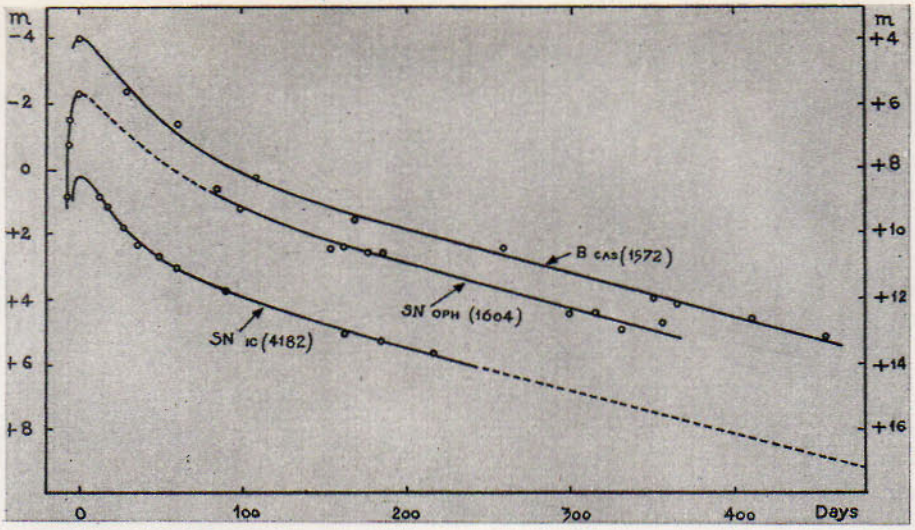


Fig. 599 — LIGHT CURVES OF THE SUPERNOVAE OF 1572 AND 1604 (MAGNITUDE SCALE ON LEFT) AND THE SUPERNOVA WHICH APPEARED IN THE EXTRAGALACTIC NEBULA IC 4182 IN 1938 (MAGNITUDE SCALE ON RIGHT).
(W. Baade, Mount Wilson Observatory)

this class: they are characterised by smaller mean periods and amplitudes, and also by more or less prolonged steady intervals at a magnitude intermediate between their maxima and minima. Now, as the Russian astronomer Parenago has shown, these stars exhibit a fairly close correlation between amplitude and mean period; and this relation appears to apply also to the recurrent novae, as can be seen from the following table:

	TYPE OF VARIABLE		NOVAE	
	Z CAMELOPARDALIS	SS CYGNI	RECURRENT	NORMAL
Mean period	10—30 days	50—100 days	10—30 years	(300—3,000 years?)
Mean amplitude (mags.)	2—3	3—5	6—8	10—13

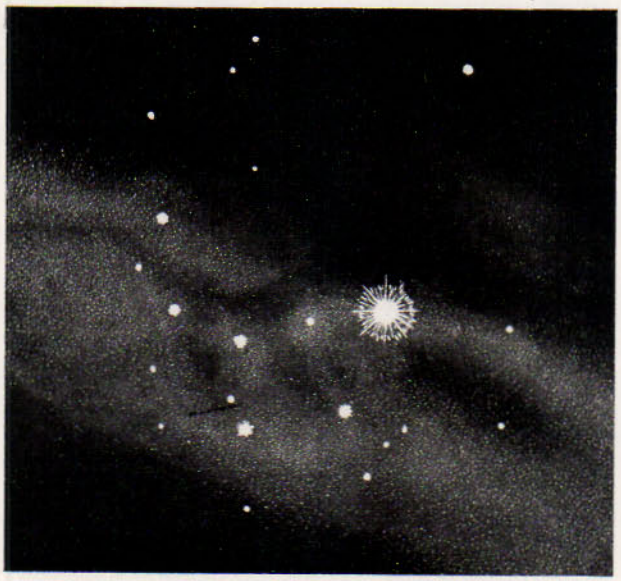


Fig. 600 — THE NOVA OF 1572, whose appearance in Cassiopeia must have been something like that shown above, caused widespread alarm and despondency.

Extrapolating from this correlation, the mean amplitude of normal novae (10—13 magnitudes) would correspond to periods ranging from several centuries to several millennia, so that it is hardly surprising that they have been observed only once. However that may be, it now appears certain that novae are no more than extreme manifestations of the internal causes leading to variation of a star's brightness, whose general effects are to be seen in the various categories of variables which we have been describing. But it is one thing to unravel the phenomena and quite another to expose their fundamental cause. This is in fact hardly understood at all, and we are still awaiting the appearance of a really satisfactory theory.

We can at least be confident that a nova is not, as was once supposed, the mutual annihilation of two faint or dark stars colliding with one another by accident, since the fantastic emptiness of interstellar space would make the chance of such encounters so small as to be negligible. Nor can the passage of a star through a cloud of diffuse interstellar matter be seriously considered. The origin of the nova-phenomenon must without doubt be looked for within the star itself, even though the precise mechanism of the outburst is at present still unknown. In Chapters 15 and 18 we shall consider the possible causes of stellar instability such as might lead to explosions of the sort that from time to time produce novae.

Supernovae. — We now come to a class of variable which has only been differentiated from that of normal novae during the last forty years or so. They share certain characteristics with novae — such as the suddenness of their appearance and the manner of their subsequent fading — but at the same time differ so markedly from them in other respects (their prodigious luminosity at maximum, their very wide amplitudes, their extremely sparse spatial distribution, and their highly individual spectroscopic characteristics) that they clearly merit inclusion in a class of their own. During the past 25 years the American astronomers, with the powerful instruments at their disposal, have studied these objects intensively.

In order to discuss the supernovae it will be necessary to anticipate some of the topics dealt with in later chapters. Stars are not scattered at random throughout all space, but are gathered into vast and widely separated stellar systems, each containing several thousand million suns.

Fig. 601 — THE CRAB NEBULA, M1, PRODUCED BY THE EXPLOSION OF A SUPERNOVA IN A.D. 1054. Photograph taken in the red Hα light of hydrogen and wavelengths λ 6518, 6584 of ionised nitrogen by W. Baade with the Mount Wilson Observatory 60-inch telescope.

That which contains our own Sun is known as the Galaxy; beyond its boundaries are the millions of other star systems, or galaxies, which populate the depths of space as far as our instruments can reach; some of these appear to us in the form of beautiful spiral nebulae, which large modern instruments have succeeded in resolving into clouds of stars.

Careful examination of photographs of one of the nearest of these extragalactic nebulae — the great spiral nebula in Andromeda — has revealed a large number of novae (more than a hundred during the years 1917—32, for example). They seem to be in every way similar to the novae of our own Galaxy, but owing to the distance of the Andromeda nebula (some 1,500,000 lightyears) their mean apparent magnitude at maximum is only about 16; this corresponds to an absolute magnitude of —6, which is altogether comparable with the luminosities of those novae which we have been considering as 'normal'.

Now one nova (S Andromedae) had already been observed in this nebula, close to the nucleus, as long ago as August 1885. After reaching a maximum of magnitude 6 it faded rapidly, just like an ordinary nova, and by 1887 had sunk out of sight (magnitude 16). For close on 30 years no one gave it another thought, but after G. W. Ritchey's discovery of faint novae in the Andromeda nebula in 1917, and the gradual realisation of their kinship with normal novae in our own Galaxy, it became evident that the nova of 1885 was something quite outside the common run of stars — even of novae. It could not, in fact, have been a normal galactic nova, for the amplitudes of these hardly exceed 13 magnitudes, as we have just seen; having attained magnitude 6 at maximum, it should therefore now be an object of magnitude 19, in which case it ought to be visible: its position was well determined by the photographs taken at the time of its outburst, and modern telescopes can record stars down to magnitude 21. But if it was associated with the nebula it represented a type of nova of altogether astounding brilliance, thousands of times greater than the already prodigious brightness of normal novae; its absolute magnitude must have been of the order of —14 or —15, which far and away exceeds that of every other known type of star.

A search of other photographs taken at the great American observatories led to the discovery of further novae of this type, and the objective reality of these unimaginably violent explosions could no longer be doubted. From that date their systematic investigation was a matter of course.

Fig. 602 — THE CRAB NEBULA PHOTOGRAPHED IN INFRA-RED LIGHT (λ 7200—8400). The filamentous structure shown in the previous picture is now almost completely invisible *(Photo, W. Baade, Mount Wilson Observatory)*

F. Zwicky has made a special study of this field, and during the years following 1936 he succeeded — with the help of a special wide-field telescope at Mount Palomar Observatory (see Chapter 19) — in observing a sufficient number of supernovae to establish the general characteristics of this type of star, several of which were also studied spectroscopically in great detail at Mount Wilson Observatory. The following table lists the brightest and best observed supernovae, together with the extragalactic nebulae in which they appeared.

PRINCIPAL EXTRAGALACTIC SUPERNOVAE OBSERVED BETWEEN 1885 AND 1940 (after C. Bertaud)

YEAR	NEBULA			SUPERNOVA	
	NUMBER	TYPE	APPARENT PHOTOGRAPHIC MAGNITUDE	PHOTOGRAPHIC MAGNITUDE AT MAXIMUM	
				APPARENT	ABSOLUTE
1885	NGC 224	*Sb*	4.5	7±	−15±
1895	NGC 4424	SB*b*	12.5	(11.1)	−15.6
1895	NGC 5253	*Sa*?	10.8	8.0	—
1901	NGC 4321	*Sc*	10.6	(11.9)	−14.8
1915	NGC 4527	*Sc*	11.3	(13.0)	−13.7
1917	NGC 6946	*Sc*	11.1	(12.9)	−12.4
1919	NGC 4486	E0	10.1	(12.0)	−14.7
1926	HGC 4303	SB*c*	10.4	(12.8)	−13.9
1936	NGC 4273	*Sc*	12.4	14.4	−12.3
1937	IC 4182	*Sc*	13.5	8.2	−16.6
1937	NGC 1003	*Sc*	13.1	12.8	−14.0
1939	NGC 4621	E5	11.4	(12±)	−13±
1940	NGC 4725	*Sb*	10.8	(12.8)	−13.6

The fundamental characteristics which emerged from these studies were: first, their absolute magnitudes of from −11 to −16, which immediately placed them in a different category from normal novae; and secondly, their spectroscopic features which are quite unlike those of any other star, and of novae in particular. Let us consider each of these in turn.

The absolute magnitudes indicate a luminosity of an order that one would have thought quite unattainable by a single star. At maxima, the supernovae are from 10 to 100 *million* times more luminous than the Sun; in fact, they appear roughly as bright as the nebulae containing them, and we know these to be gigantic stellar systems containing thousands of millions of stars. It has been calculated from the light curves that during the few months that it is visible, a typical supernova radiates as much energy as the Sun dispenses in several million years.

Language is inadequate to describe such a flood of energy: the best we can do is cite examples based upon the above figures. Thus a supernova appearing in our own Galaxy at a distance of 10 parsecs (33 lightyears) would have an apparent magnitude equal to its absolute magnitude; in other words, although it would appear as a dimensionless stellar point, its brightness at maximum would be 10 times that of the full Moon, and it would turn the night into day.

The spectra of the supernovae, which resemble those of no other type of star, have presented astronomers with a nice set of problems on which they are still exercising their ingenuity. Without going into technical details let us simply remark that S. and C. Gaposhkin have suggested a likely interpretation, according to which the spectra are produced by two common elements (iron and hydrogen) and are displaced by the high radial velocities (of the order of 3000 to 6000 m.p.s.) with which the incandescent gases are expelled from the explosion centre. The spectroscopic changes proceed with such extreme regularity, and show such perfect synchronisation with the light variations, that if a supernova is discovered when already in its decline, an examination of its spectrum will at once tell us the exact number of days that have elapsed since its maximum.

The light curves of the supernovae in fact exhibit much less diversity than those of novae. However, Baade, at Mount Wilson, has been able to distinguish two main types of supernova, the distinction resting on spectra and maximum absolute magnitudes. Type I supernovae have sharp maxima, followed by a rapid decline which slows down steadily; their absolute magnitudes are high (−14 to −16); and their spectra are of the type described above. Type II supernovae have broader maxima; they fade rapidly at first, then slow down, and finally accelerate again some 85 days after maximum; their light curves are therefore quite characteristic; their absolute magnitudes are lower than those of Type I supernovae (−11 to −13); and their spectra show analogies with those of normal novae. Thus they appear to constitute a sort of half-way house between Type I supernovae and ordinary novae.

For this reason it has been suggested that supernovae merely represent an extreme and rarely occurring case among normal novae. This is not certain, however, for whereas the latter are concentrated strongly about the central planes of the star systems in which they occur, supernovae are found anywhere throughout them, and perhaps even show a slight preference for the outlying regions.

The number of supernovae that Zwicky detected in the course of a few years should not mislead us as to their true numbers. His investigation covered some thousands of galaxies, and when this and the time spent on the survey are taken into account it would appear that on the average only 2 supernovae occur per 1,000 years in any one galaxy. This is only a mean figure, and cases are known where two supernovae have appeared in rapid succession (i.e. within less than 20 years) in the same galaxy — for example, in 1901 and 1914 in NGC 4321.

Supernovae are thus some 10,000 times rarer than normal novae, and this extreme rarity shows that we have only a slender chance of observing one in our own Galaxy during the course of our brief lifetime. None the less, a search of the past annals of astronomy for exceptionally brilliant temporary stars has uncovered the records of three which can confidently be identified as supernovae. The most remarkable was the celebrated nova which appeared early in October 1572 and was carefully observed by the Danish astronomer Tycho Brahe, who has left us a detailed account of its behaviour. At maximum (Fig.600) it outshone Venus, and was visible in broad daylight; but this phase passed quickly, and by the end of the year it had sunk to about magnitude 0. Six months later it had passed magnitude 2, and early in 1574 it sank below the sixth magnitude and was no longer visible to the naked eye. Baade, from Tycho Brahe's records, has reconstructed the light curve, which is characteristic of a Type I supernova (Fig. 599).

The position of Tycho's star is known with reasonable accuracy from the contemporary records, and in recent years a search was made in an endeavour to identify it — in whatever modest guise it may now have. But even the great Mount Wilson reflectors were unable to find any trace of it, and it is now certainly fainter than magnitude 18. Its amplitude must therefore have been at least 22 magnitudes, far outstripping the 14 magnitudes which is the greatest amplitude of normal novae. Its brightness at maximum and minimum must have differed by a factor of at least 10^9. From this it is also deduced that its present luminosity is very low, corresponding to an absolute

magnitude of at least 6 — assuming, of course, that the star was not completely vaporised by the heat of the explosion, an eventuality which we shall see later is in fact not very likely.

Thirty-two years later, in 1604, another outstandingly bright star made its appearance — this time in Ophiuchus — and was studiously observed by Kepler. At maximum it was brighter than Jupiter, and the contemporary records allow us to reconstruct its light curve; once again, this is precisely similar to that of a Type I supernova such as that observed in 1937 in the nebula IC 4182 (Fig. 599). There is thus no doubt that this outburst also was a supernova situated in our Galaxy, for had it been a normal nova (amplitude not exceeding 14 magnitudes) its present apparent magnitude would be at most 11 or 12 and it would be readily visible. This, however, is not the case, although photographs taken in red light at Mount Wilson have revealed a small nebulosity close to its 1604 position, and this may well be the gaseous envelope which was expelled at the time of its outburst.

Corroboration of this conjecture is provided by the beautiful and strange Crab Nebula (Figs. 601 and 602; so called after its vague resemblance to a crab), discovered in 1764 by Messier, from his observatory at the top of the Tower of Cluny in Paris. This object has intrigued astronomers for a long time, since it is unique of its kind and cannot be classed with any other known nebula. Furthermore, the study of its spectrum and photographic measurements made at intervals of half a century or so have shown that it is expanding rapidly — at a rate of something like 600 m.p.s. Extrapolating backwards in time, it is found that this expansion must have started about 800 or 900 years ago. Now during the course of an examination of ancient astronomical records which he undertook some 25 years ago, the Swedish astronomer Lundmark discovered that in A.D. 1054 the Chinese and Japanese astronomers observed a bright nova in exactly the position now occupied by the Crab Nebula. Its brightness was equal to that of Jupiter (mag. −2). Since the distance of the nebula is known (5,000 to 6,000 light-years), the absolute magnitude of the star can be calculated from its apparent magnitude of −2: it must have been −13. Clearly it was a supernova, and the nebula was its offspring. Two faint stars (magnitude 16) can even be detected at the heart of the nebula, and one of these appears to be the remnant of the supernova.

Thus three supernovae have appeared in our Galaxy during the course of the past millenium, which corresponds fairly closely with their observed frequency in external galaxies. Perhaps in the not so distant future we shall have the opportunity of observing a fourth supernova in our vicinity.

It has to be admitted that nothing certain is yet known of the origins of supernovae. The sheer multiplicity of the explanatory theories is enough to prove this; besides, we can hardly expect to have found, thus soon, the explanation of a phenomenon so unprecedented, so unexpected, and so recently subjected to adequate study.

The field of theoretical explanations can only be briefly glanced at. According to Gaposchkin it is in fact only right and proper to call these objects *super*-novae — i.e. novae of exceptional violence — and the discovery of the intermediate Type II supernovae supports this view. Zwicky believes that they are stars whose core suddenly collapses (for reasons which we shall discuss later); the gravitation potential energy, thus suddenly released, takes the form of a flood of radiation whose colossal pressure blows out the star's external layers.

Recently E. Schatzman, of the Institut d'Astrophysique de Paris, has suggested that the supernovae may be stars under very special physical conditions (which would explain their rarity), a region of whose interior by chance becomes extremely unstable and sets up a shock wave which is propagated at enormous velocity towards the surface, where it dislodges and expels a small part of the star's total mass.

But, once again, we must await the accumulation of more numerous and comprehensive observations before we can arrive at any definite conclusions about the causes of these fantastic phenomena — surely the most grandiose and awe-inspiring that a star can undergo or that we are privileged to witness.

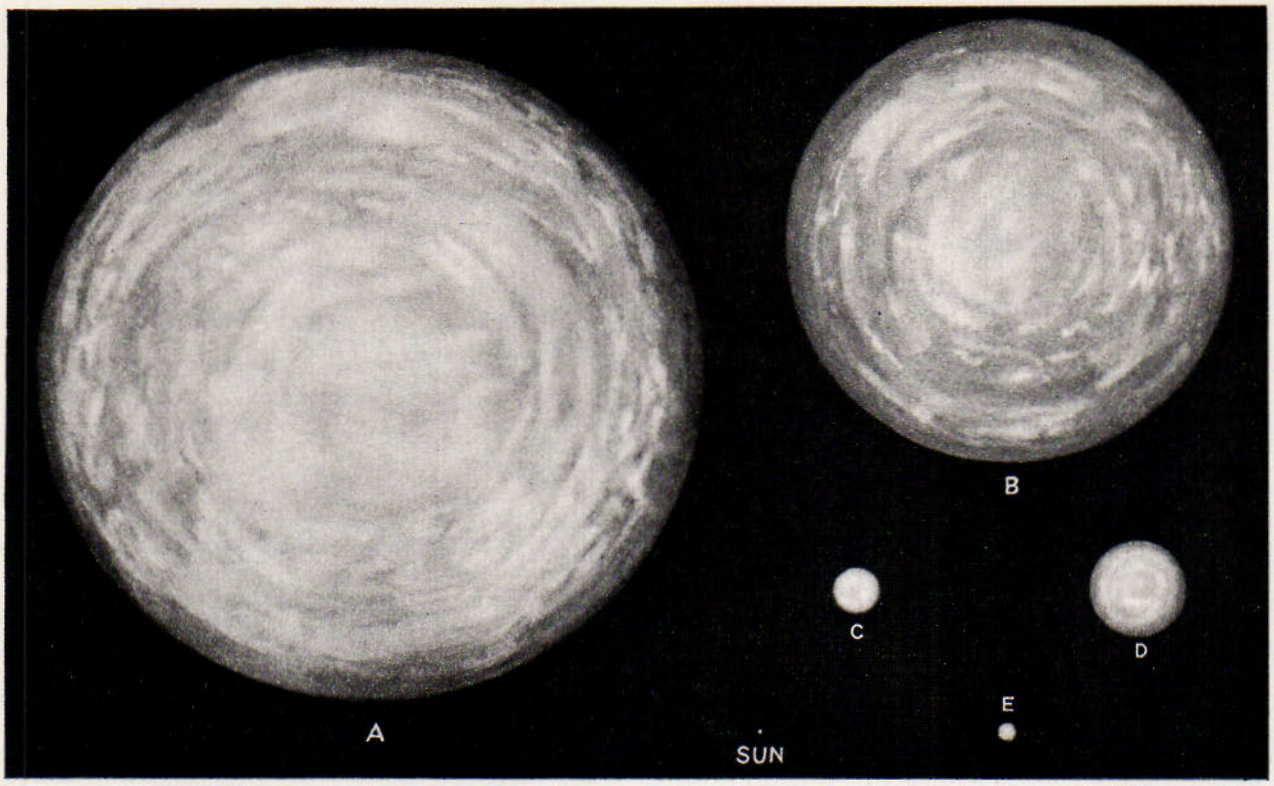


Fig. 603 — SOME GIANT STARS. A, ε Aurigae; B, β Pegasi; C, α Aurigae; D, α Boötis; E, Y Cygni. On the same scale the Sun would be rather larger than this full stop.

CHAPTER 14

THE STARS AS SUNS

In the preceding chapters we have made the acquaintance of the universe of the stars, but we were still considering them as infinitely remote objects — mere dimensionless points of light, some bright, some faint. Now, thanks to the power of modern instruments and observational methods, we can strip away this illusory appearance and gain some idea of the facts and figures that underlie the centuries-old dictum, whose full significance has only been realised in our own times: 'the stars are other Suns'.

The luminosities of the stars. — The apparent brightness of a star is obviously determined by two independent factors — its intrinsic luminosity and its distance from the terrestrial observer. It is a familiar fact of everyday experience that the observed brightness of an electric light bulb, for instance, varies according to the distance from which it is observed: its brightness in fact diminishes in proportion with the square of its distance. Thus if one bulb is viewed at 10 times the distance of another (of the same intrinsic brightness) it will appear to be 100 times fainter. Conversely, if two bulbs appear to be equally bright although one is actually 100 times more luminous than the other, then the latter must be 10 times nearer than the former.

Thus the fact of two stars having the same apparent magnitude tells us nothing about their luminosities. Since both the distances and the luminosities of the stars vary over an enormously wide range, we cannot usefully compare one star with another unless the effects of their different distances are first eliminated. To do this, it is calculated what their apparent brightnesses would be if they were both at the same distance from the observer; for convenience this standard distance is taken as 10 parsecs (about 33 lightyears), and the apparent magnitude that a star would have at this distance from the observer is called its absolute magnitude (denoted by M to distinguish it from the apparent magnitude, m). Reduced by calculation (based on the inverse square law) to this common basis for comparison, stars of every grade of apparent brightness assume their true rank in the stellar hierarchy.

Let us take the case of the Sun, for example. Owing to its proximity it appears blindingly brilliant: its apparent magnitude is in fact -27. But if it were removed to the standard distance of 10 parsecs it would become a faint star of magnitude 5 — its absolute magnitude, in other words, is $+5$. On the other hand a star which is hardly visible in the largest telescope but which is also very remote might have an absolute magnitude of, say -1 — i.e. if brought in to the standard distance it would appear a brilliant object. It might be added that our nearest stellar neighbour, Proxima Centauri, has an apparent magnitude of 11 (requiring a telescope to show it), and

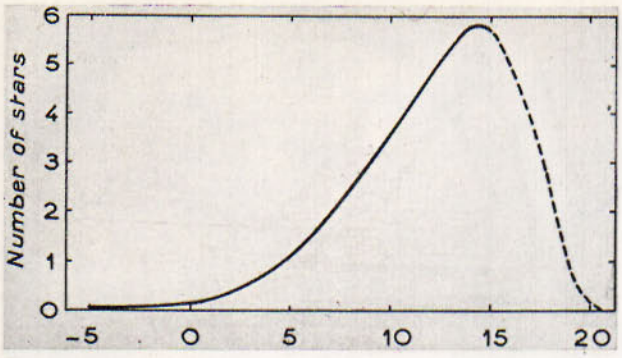


Fig. 604 — THE LUMINOSITY FUNCTION RELATES THE NUMBERS OF THE STARS TO THEIR ABSOLUTE MAGNITUDES

by reason of its low absolute magnitude ($M = +15$) it would require a very powerful instrument to show it at a distance of 33 lightyears.

Thus the differing apparent brightness of the stars does not spring only from their greater or smaller distances from the observer; this is simply a factor modifying, in one direction or the other, the appearance of suns which also differ very widely among themselves as regards both their dimensions and the intensity of their radiation.

When we come to discuss these stellar characteristics in greater detail we shall find that they vary over an enormously wide range. Speaking generally, absolute magnitudes vary from -5 for the supergiants to $+15$ for the red dwarfs; whence it follows that the former radiate 10^8 times as much energy as the latter in a given time. The Sun occupies an intermediate position between these two extremes of luminosity: its absolute magnitude of $+5$ shows that it radiates 10^4 times the energy of an average dwarf and 10^{-4} that of an average supergiant.

The manner in which the stars are distributed on the absolute magnitude scale is revealed by what is known as their luminosity function (Fig. 604), whose exact determination constitutes a problem not yet completely solved, for although it is relatively easy to assess the numbers of the intrinsically very brilliant stars, which are visible at great distances, it is much more difficult to differentiate, among the myriads of apparently faint stars, between those that are intrinsically faint and those that appear faint merely because of their great distances. The investigation of stellar proper motions, however, gives us a weapon for attacking this problem, since large angular displacements must, on the average, indicate comparative proximity. In any case, the currently accepted luminosity function curve shows clearly enough a strong concentration about absolute magnitudes $+14$ to $+15$ — i.e. 10^4 times fainter than the Sun, which therefore cuts quite a good figure in the total assemblage of the stars.

It will also be seen from the diagram that stars even 100 times more luminous than the Sun are very rare, while those with luminosities of from 10^3 to 10^4 times the Sun's are altogether exceptional.

It should be mentioned that the validity of these figures is restricted to the neighbourhood of the Sun, for it is only here that we can hope to derive correct statistics for the intrinsically fainter stars. It is by no means certain that the same proportions between numbers and absolute magnitudes apply universally; indeed, we shall see that in certain cases there are strong indications that they are far from being valid throughout all space.

For this reason it is worth enquiring what exactly we mean by the 'neighbourhood' of the Sun, so that we can form some sort of idea of the extent of the domain covered by our observations. The table of the nearest stars (p. 310) shows that of these stars four figure in the list of stars of the first apparent magnitude (Sirius, Altair, Procyon and α Centauri) and only one, Sirius, is of the first absolute magnitude. This underlines the fundamental nature of the distinction between absolute and apparent magnitude. In view of the widely disparate luminosities of the stars, therefore, it does not mean a great deal to speak of 'the' distance of the stars of such or such an apparent magnitude: so much depends on their absolute magnitudes. An example — which at the same time gives us some inkling of the spatial scope of our explorations — will help to make this clear.

The distance of a magnitude 11 star, just visible in a telescope of about $\frac{1}{4}$-in. aperture, will be 4 lightyears if its absolute magnitude is $+15$ (as is the case of α Centauri C, or Proxima Centauri, the nearest known star), rising to 40 lightyears if its absolute magnitude is $+10$, 400 lightyears if it is $+5$ (that of the Sun), and 4,000 lightyears for a highly luminous star of $M = -5$ (such as Rigel). In this last case, however, a correction would have to be applied to eliminate the effect of diffuse interstellar matter, whose absorption of the starlight passing through it becomes an important factor when distances of this size are involved (see Chapter 16). Nevertheless, we can see from this that the volume of space within which we can at present observe about 10^9 stars extends for at least several thousand lightyears all around us.

The sizes of the stars. — Just as the stars vary very widely among themselves as regards their luminosities, so they are very far from being all of the same, or even similar, size. Although this has been suspected for a long time, it is only comparatively recently that exact information has been obtained.

The real dimensions of the stars must remain completely unknown so long as we are restricted to simple telescopic observation, for even in the most powerful instruments the stars remain mere points of light, without any appreciable size. The high-magnification examination of a star with any telescope (even a small one) might at first sight appear to contradict this statement, since it will be seen, not as a point, but as a minute disc surrounded by a number of concentric and very faintly coloured bright rings. This, however, is a purely optical effect due to the diffraction of light. Similarly, the photographic images of stars are circular spots of more or less sensible diamater, due to the diffusion of light in the photographic emulsion. This is a function of the brightness of the image and the length of the exposure: the brighter the star, or the longer the exposure, the larger will be the photographic image. In this field of physical optics certain modern developments have opened the door to the direct measurement of the true linear diameters of the stars.

For the moment, however, we shall concern ourselves with the indirect method of determining stellar size, the basis of which is the discovery of a star's absolute magnitude. This, clearly, will depend on (*a*) the radiant energy emitted per unit area of its surface, which in turn is a function of the star's effective temperature, (*b*) the size of this surface, and hence the radius of the star. The relation between these quantities is used to deduce the star's diameter from its absolute magnitude and temperature; at the same time its distance is derived by comparing its observed apparent magnitude and its absolute magnitude as deduced from its spectrum. The accuracy of the final result thus depends upon that of various intermediate operations, and these are often rather difficult to undertake with a satisfactory margin of error. Nevertheless, the method has at least revealed the order of stellar size, and its results are sufficiently near to the truth.

Moreover, these results have on the whole been confirmed by the interferometric method (Chapter 19) —

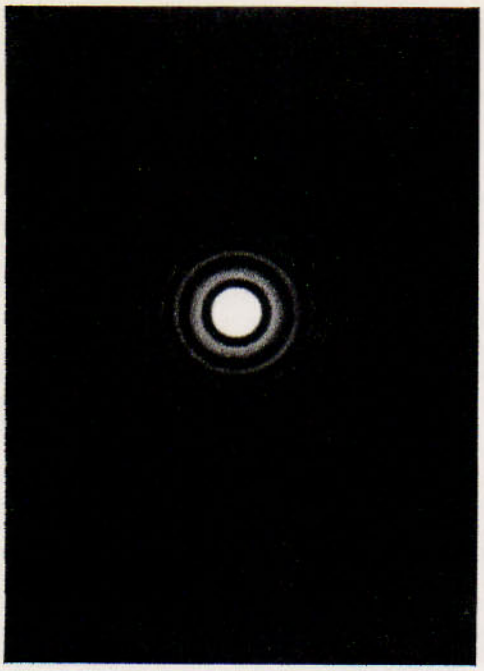


Fig. 605 — When observed with high magnifications in even the largest telescopes, a star shows nothing but a spurious disc surrounded by a number of diffraction rings resulting from the undulatory nature of light.

developed by the American physicist Michelson — which is capable of yielding direct measurements of some of the larger stars: for a given star, the results of the two methods are practically identical. When the indirect method is the only one that can be employed — as is true in the vast majority of cases — we can therefore accept its findings with confidence.

A third method of determining the linear sizes of the stars consists of the analysis of the light curves and radial velocities of eclipsing variables; this allows the dimensions of the two stars to be calculated, often with considerable precision, for it will be obvious that the relation of the duration of eclipse to the whole period depends, among other things, on these very dimensions: the larger the eclipsing body, the longer the duration of eclipse.

Let us now review the results that have been obtained by these various methods. So far as our present knowledge goes, the record for size is held by the primary component of the eclipsing binary ε Aurigae, whose diameter is approximately 2,000 times that of the Sun.* The runner-up is VV Cephei, another double, whose primary is estimated to be 1,200 times the diameter of the Sun. These are followed by a series of smaller but still gigantic stars. In addition to them there are many stars which are comparable with the Sun in size, and others again by comparison with which the Sun is itself a giant. The table on the next page lists, in order of decreasing size, some of the stars whose dimensions have been determined with satisfactory accuracy. The data of the largest reveal the good agreement that exists between the direct method (calculated diameters) and the interferometric method (measured diameters).

The enormous diversity of stellar size will be noted; in particular, that the last 4 stars are quite microscopic compared with the others — they are, in fact, smaller than the major planets of the Solar System, while the last 2 are even smaller than the Earth. We are already familiar with the sizes of some of the stars listed, but not with those of the giants and supergiants. In order properly to

* The diameters of the components of another eclipsing binary, V 381 Scorpii, have been determined (though less certainly) as 1,500 and 3,000 times that of the Sun.

Fig. 606 — Stars appear on photographs as circles of diffusion (which are very much larger than the diffraction image) whose size is proportional to their apparent brightness. The brightest images of all are surrounded by a photographic halo. The square inscribed on this photograph represents an area 100 times larger than that of the plate above.

MEASURED AND CALCULATED DIAMETERS IF REPRESENTATIVE STARS

STAR	SPECTRO-SCOPIC TYPE	MEASURED DIAMETER (⊙ = 1)	CALCULATED DIAMETER (⊙ = 1)	
ε Aurigae B	K5	2,000		
VV Cephei A....	M2	1,200		
o Ceti	M6e	*400*		Supergiants
α Orionis	M2	*300—400***	400	
ξ Aurigae A	K5	300		
α Scorpii	M1	*300*	320	
ε Aurigae A ....*	F5	200		
β Pegasi	M5	*110*	130	
α Tauri.....	K5	*36*	57	
V 380 Cygni A ..	B2	29		Giants
α Boötis	K0	*23*	26	
α Aurigae	(G0)		16	
γ Cygni A, B ...	O9	5.9		
β Persei A	B8	3.1		
β Aurigae A	A0	2.8		
Procyon A	F5		1.7	
α Centauri E	G4		1.2	
Sun.....	G2	1.0	1.0	Main Sequence Dwarfs
W Ursae Majoris A ...	F8	0.9		
70 Ophiuchi A ..	K1		0.9	
70 Ophiuchi B...	K5		0.7	
Krüger 60 A	M4		0.5	
Sirius B	A5		0.02	
40 Eridani B	A		0.02	White Dwarfs
van Maanen 2 ...	G		0.006	
Wolf 457	(a)		0.003	

* The italicised values are derived from Pease's interferometric measurements modified to take account of the best recent parallax measurements; the remainder are derived from the light curves of eclipsing binaries.

** Variable diameter.

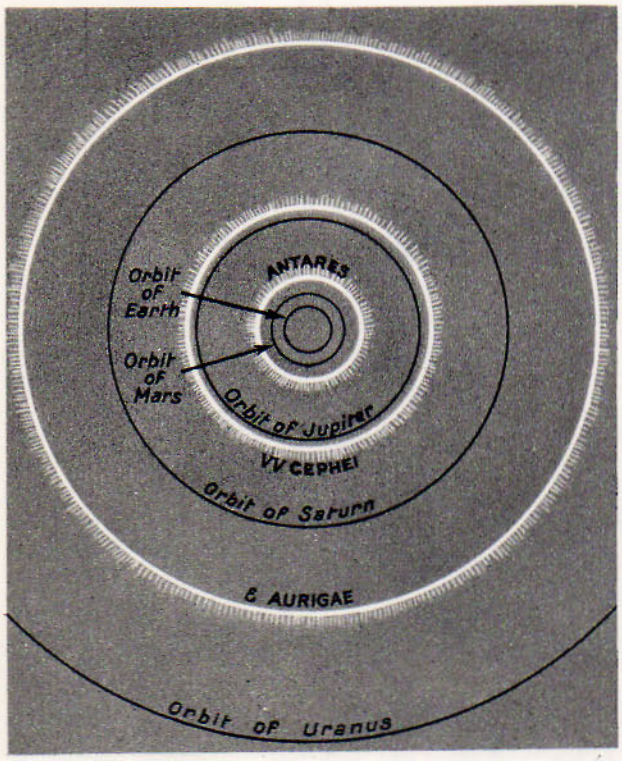


Fig. 607 — DIAMETERS OF SOME SUPERGIANTS COMPARED WITH PLANETARY ORBITS

appreciate the staggering sizes of these it is helpful to compare some of them with the scale of the Solar System, supposing that they have replaced the Sun: then the gaseous mass of β Pegasi would engulf the orbit of Venus, the photospheres of Betelgeuse and Mira would lie nearly at the orbit of Mars, while the enormous spheres of α Herculis and Antares would reach beyond it. These vast stars would in turn be swallowed up by the still more colossal globes of VV Cephei, ε Aurigae and V 381 Scorpii, the first of which would occupy the whole of the Solar System as far as the orbit of Jupiter, while the latter two would reach or pass that of Saturn.

One is left stupefied in the face of such incredibly varied dimensions: compared with some the Sun is a dwarf, with others a giant. In fact the Sun's middle-class status in stellar society is not too undistinguished when one bears in mind the fact that the giants constitute a sort of *élite* among the vast throngs of the smaller proletariat. Since the Sun lies roughly midway between the extremes of stellar size and luminosity it is convenient to describe these in terms of the Sun's diameter and intrinsic brightness as unity. In the table, stars whose diameters and luminosities exceed the Sun's by a factor of 100 or more are termed supergiants; those which are a few score times superior to the Sun are termed giants; those comparable with the Sun are dwarfs, the smallest of which, much fainter intrinsically than the Sun, are sometimes called red dwarfs on account of their characteristic colour; and, finally, the minute and very faint stars at the foot of the table belong to the class of white dwarfs, about which we shall have more to say later.

Surface temperature and colour of the stars. — The diversity shown by stellar luminosity and size is a reflection of the stars' physical characteristics, and these must now be described — at any rate in outline.

The intimate nature and varied physical conditions of the stars can only be investigated in detail by physical methods, of which spectroscopic analysis is the most important. This will be dealt with in Chapter 20, but some clues are furnished by even a casual visual examination — for this will at least tell us the colour of a star.

Among the brighter naked-eye stars, some are white or bluish-white (Sirius and Vega, for example), others are yellow (Arcturus, Capella), others again are orange or reddish (Aldebaran, Antares). These differences are sufficiently obvious to the eye and have therefore been recognised from the earliest times. A similar diversity of colour is encountered among the fainter, telescopic stars, but for physiological reasons is less noticeable than among the brighter ones. Photography, on the other hand, always registers these differences with perfect fidelity, thus providing a measure of the colour indices of the stars, as already explained.

As everybody knows, the colour of an incandescent source depends on its temperature. As a lump of iron is heated it will begin to glow with a dull red colour, changing progressively to a bright red, orange, yellow, and finally a blinding white; on cooling, it retraces its steps along the same series in the reverse order. The same phenomenon is to be encountered among the different forms of domestic lighting: the carbon filaments of the earliest electric bulbs incandesced at only 2,000° and gave a deep yellow light compared with the much whiter light of a tungsten filament which incandesces at 2,900°; and both these appear yellowish when compared with the blinding whiteness of an electric arc (3,900°). This is, of course, a rough guide only, since the temperatures necessary to make different elements incandesce are not all the same. It must also be remembered that what we call white light is a mixture of radiations of many different

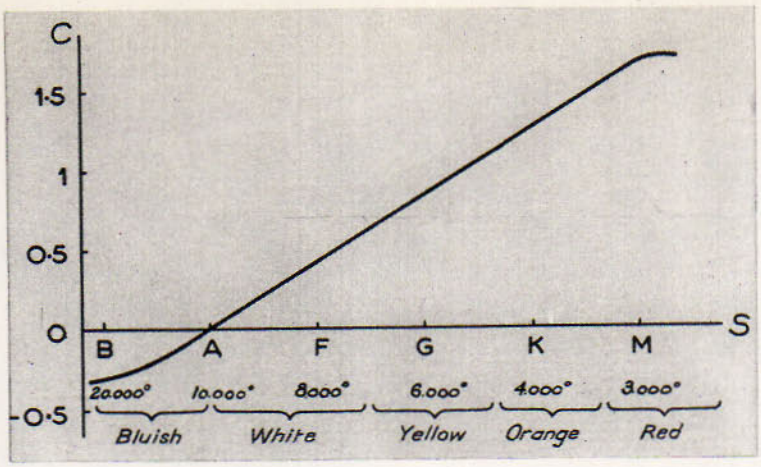


Fig. 608 — VARIATION OF STELLAR COLOUR INDEX AS A FUNCTION OF TEMPERATURE

wavelengths; the longest to which our eyes are sensitive give the subjective sensation of redness, and the shortest that of violet.

Let us now return to a consideration of the stars as sources of radiation. The temperature of a star not only determines the intensity of the radiation that it will emit per unit area of its surface, but also the relative proportions in which radiations of different wavelength occur. Thus a star with a surface temperature of 3,000° C will appear 15 times fainter than one of the same size at a temperature of 5,000°; at the same time, as we have seen above, their colours will be different.

So far as we know at present, the temperature of a star must be at least 1,500° to 2,000° C in order to emit enough visible radiation to be seen, and at such temperatures the greater part of it will lie in the red region of the spectrum. At higher temperatures it will still emit plenty of red light — more, in fact, than it did at the lower temperature — but this is masked by its still more intense radiation of yellow light, and, at higher temperatures still, of blue. Stars never exhibit pure colours, but rather appear to have a general whitish tint which is tinged to a greater or smaller extent with blue or with yellow verging to red, as the case may be; no matter what wavelength may be emitted most profusely, there is always a wide selection of others contributing to the total radiation of the star.

It is possible to measure the temperature of a star with considerable accuracy by means of sensitive photometric determinations of the proportions of the different colours that make up its integrated light, and the comparison of these with a standard lamp whose temperature has been accurately determined in the laboratory (see Chapter 20).

In this way it has been established that the surface temperatures of the red stars are about 2,000° to 3,000° C, of orange stars from 3,000° to 5,000°, of yellow from 5,000° to 8,000° (that of the Sun is 6,000°), of white from 8,000° to 12,000°, and of bluish stars from 12,000° up to 20,000°, 30,000° or perhaps even higher, though the measurement of these extreme temperatures is more difficult and the results correspondingly uncertain.

Thus even the simple appreciation of the colour of a star will give us some idea of its probable temperature. A more accurate figure will naturally be derived from its colour index, which, as we have seen, is the difference between its photographic magnitude (blue and violet light) and its visual magnitude (yellow and green). By convention, the value of the colour index is taken as zero for a certain class of white stars; from this zero point, negative values increase towards the blue and positive towards the red, as shown in Fig. 608.

Spectroscopic classification of the stars. — But the best and most certain determinations of stellar temperature are those that are based on the detailed spectroscopic analysis of their light.

A general account of the principles and applications of spectrum analysis will be found in Chapter 20, where it is explained how the dark or bright lines and bands in stellar spectra are tied up with the absorption or emission of radiant energy by atoms and molecules which can be identified with absolute certainty. We shall also see that these spectral lines constitute what is in effect an identity card of the body (simple or composite) which is producing the radiation, and that they vary according to the conditions of temperature and pressure in the source. Thus at the various temperatures of the Bunsen burner (2,000°), the electric arc (3,000°) and the electric spark (5,000° to 10,000°), the same gas will emit different sets of radiations; these, however, are in every case peculiar to itself, so that the spectrum allows us to identify at the same time the chemical constitution and the temperature of the source.

Without entering into a lengthy theoretical disquisition, let us simply record the fact that these spectroscopic changes correspond to different states of the atoms of the radiating gas, and that different sets of lines appear in turn as the temperature is raised and as its atoms, originally neutral, become increasingly highly ionised — i.e. as they lose more and more of their orbital electrons. Thus the relative intensities of the different lines of a single element in the spectrum of a star reveal the degree of ionisation of that element, and hence the temperature of the radiating layer. However, the absence of the characteristic lines of a particular element from the spectrum of a star does not necessarily imply that it does not occur in that star at all: the physical conditions may not be favourable for its emission of these wavelengths, or, alternatively, they may lie in regions of the spectrum which cannot be observed. It has already been explained that the terrestrial atmosphere absorbs certain wavelengths from the light of the stars passing through it: thus we are prevented from knowing anything about their spectroscopic characteristics within these zones of absorption. Ozone in the upper atmosphere cuts out the greater part of the abundant ultra-violet radiations of the very hot stars, while water vapour and carbon dioxide absorb many of the infra-red wavelengths which predominate in

the spectra of the coolest stars. Despite these drawbacks, fundamental data can be deduced from the spectra of the stars.

These spectra can be divided into a small number of well-defined types, which are designed by the letters shown on the horizontal axis of the colour index diagram (Fig. 608). The classification is built up of a series of small, progressive spectral differences in such a manner that the majority of the stars are arranged in order of decreasing temperature.

An early attempt to classify the stars along these lines was made by Father Secchi in 1867. His spectroscopic observations led him to distinguish three main classes of spectra: white stars (Type I), yellow (Type II) and red (Type III), in order of decreasing temperature. Since Secchi's day such progress has been made in spectroscopic studies that the adoption of an increasingly detailed system of classification has been called for.

The system universally employed at the present time was developed during the early years of this century as a result of the herculean investigations of stellar spectra undertaken at Harvard College Observatory, which culminated in the *Henry Draper Catalogue* of over 225,000 stellar spectra. These were distributed among ten principal classes in order of decreasing temperature, so that the members of each spectroscopic type have approximately the same temperature. These types were originally designated by letters in alphabetical order, but owing to the suppression of some and the insertion of others — necessitated by our rapidly expanding knowledge of stellar spectroscopy — the series now runs:

O — B — A — F — G — K — M — R — N — S.

A useful mnemonic for this is 'Oh Be A Fine Girl, Kiss Me Right Now Sweetheart'.

Since the spectral modifications are continuous throughout each type, each of these has been further subdivided. Thus type O, which comprises the hottest stars, contains ten sub-types distinguished by a small letter (from *a* to *e*) or a number from 1 to 5). The types from B to M are decimally subdivided from 0 to 9; for example, a type G star whose spectrum is in many respects similar to those of the following type, K, might be allocated to sub-type G8. Finally, the last three types R, N, and S, whose characteristics are generally similar to those of K and M, are considered as side branches from the main series; the R—N branch being an offshoot from type K, and type S following type M. From one end of this series to the other, the spectroscopic characteristics change continuously and smoothly, the lines increasing or decreasing regularly and by insensible stages, from the hottest stars to the coolest.

The principal characteristics of each type are as follows:

Type O: The so-called ionised helium stars, whose spectra are characterised by the lines of ionised helium and those (doubly or trebly ionised) of oxygen and nitrogen. This state of ionisation indicates a very high temperature, of the order of 30,000° — 40,000° C or more. The sub-types O*a*, O*f* and O*c*, with their wide emission bands, are known as Wolf-Rayet stars (after the two French astronomers who discovered them, nearly a century ago); the temperatures of the Wolf-Rayets must be fantastically high (50,000° — 100,000° C possibly), and a special type (W) has recently been created to accommodate them. Type O stars are by no means numerous, hardly a hundred being known; of these, only three can be seen easily with the naked eye — the magnitude 2 γ Velorum (type O*a*), ζ Puppis (type O*a*) and ζ Orions (O*e*5).

Type B: The neutral helium stars; their temperatures of from 25,000° to 12,000° C are too low to ionise the helium atoms, though the lines of ionised oxygen and nitrogen are still to be seen; the intensity of these decreases rapidly as does that of the helium lines, from B5 onwards.

On the other hand the characteristic series of hydrogen lines and the wide hydrogen absorptions in the ultra-violet, which first make their appearance towards the end of type O, become progressively more marked throughout type B, and are the most noticeable feature of its later sub-types. The spectra of many early B type stars show emission lines of hydrogen (i.e. lines which are brighter than the continuous background of the spectrum); this feature, indicated by the suffix *e*, is attributed to the existence of very extensive and rarefied atmospheres surrounding these stars. γ Cassiopeiae (type BO*e*) is an example.

B type spectra sometimes contain two more or less intense lines of ionised calcium, but it is known that their origin does not lie in the reversing layer of the star itself, but in interstellar space, where atoms of this gas have been detected (see Chapter 16). Stars exhibiting this feature are distinguished by the index *k*; all other anomalies are indicated by the suffix *p*.

B type stars figure prominently in our star catalogues. But, as we shall see later, this is due mainly to their high luminosities, which allow them to be seen from great distances. They are, in fact, of comparatively infrequent occurrence in space, though it is often to be noted that they appear to favour particular regions: thus Orion contains a large group, which includes the brightest member of the type, β Orionis (apparent magnitude 0.4, type B8*p*), γ (magnitude 1.7, type B2) and δ (magnitude 1.7, type BO) and others.

Type A: Also known as hydrogen stars, owing to the importance of the lines of this element in their spectra; they attain their maximum intensity in A0 and A1, after which they become progressively weaker. The temperatures of 12,000° to 8,000° C result in the disappearance of the helium lines and the appearance of those of metals and of ionised calcium, which become rather intense in the later sub-types.

Type A stars appear to be very numerous, both because of their great absolute brightness and because of their real abundance in space. The finest examples of this type are Sirius (α Canis Majoris) and Vega (α Lyrae), both A0; other bright A type stars are Deneb (α Cagni), A2, and Altair (α Aquilae), A5.

Type F: Sometimes known as ionised calcium stars; the intensity of the two lines of this element in the extreme violet and ultra-violet increases progressively from F0 to F9, finally surpassing that of the hydrogen lines, whose intensity continues to decrease; at the same time fine and very numerous metallic lines develop steadily. The temperatures of these stars lie between about 8,000° and 6,000°. F type stars are less widely represented — at any rate in the catalogues — than those of neighbouring types. The brightest examples are Canopus (α Carinae) F0, and Procyon (α Canis Minoris) F3; Polaris (α Ursae Minoris) also belongs to this type (F8).

Type G: Known as solar type stars owing to the Sun's membership of the class. The main G type characteristics are the enormous development of the lines of ionised calcium (whose intensity approaches its maximum in the later sub-types) and the abundance of the metallic lines, more than 12,000 of which have been catalogued in the solar spectrum; the best represented metal is iron, which contributes more than 50% of the total number of metallic lines.

In this class the lines of hydrogen are no longer conspicuous, and continue to decrease in intensity, while those of neutral calcium make their appearance. The temperatures (6,000° — 4,000° C) of G stars are in fact too low to produce ionisation in the majority of the elements.

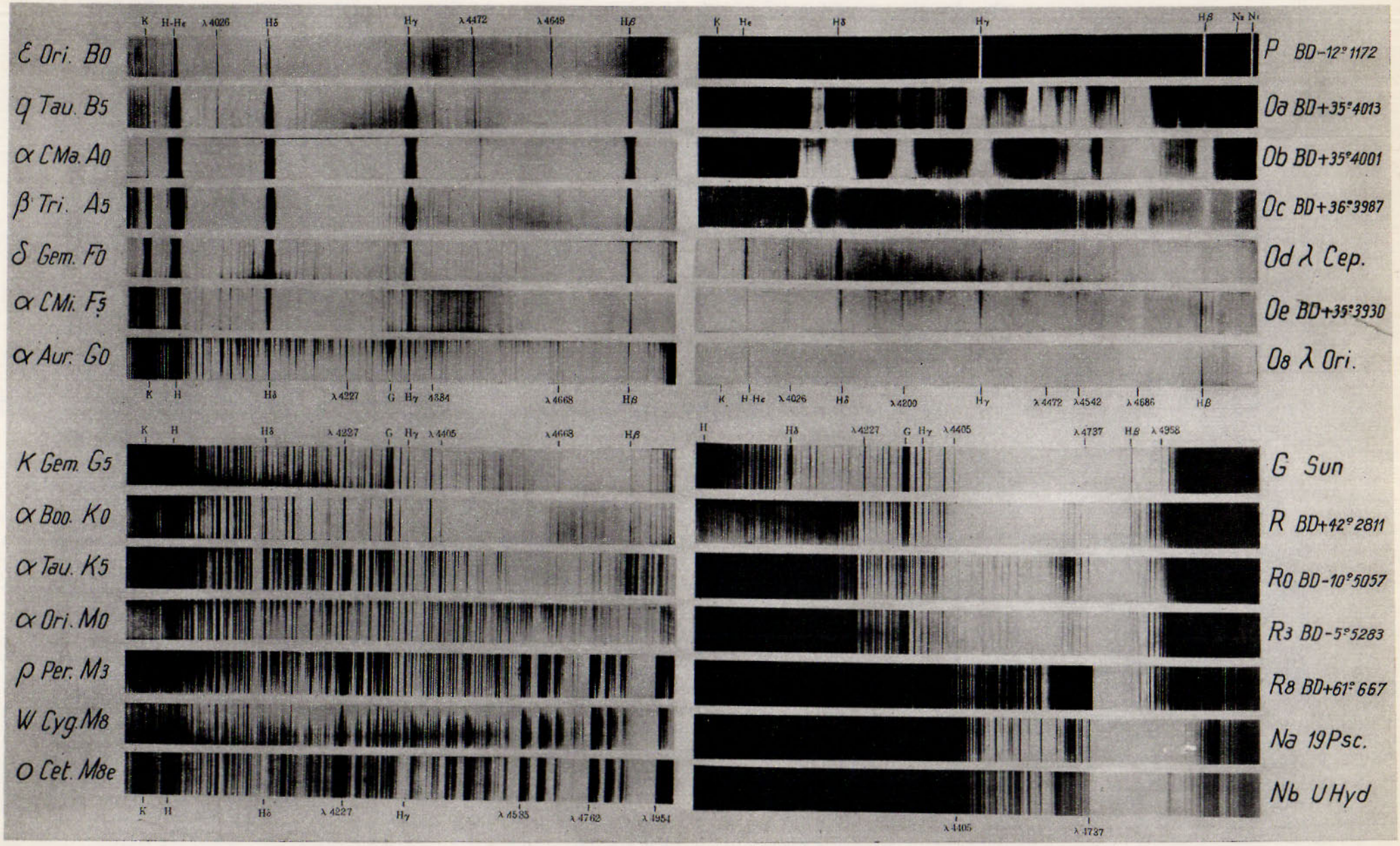


Fig. 609 — STELLAR SPECTRA IN THE BLUE-VIOLET REGION. Left: the series B, A, F, G, K, M — the commonest spectroscopic types. Right: rarer types; above, types P (nebular), O*a*, O*b*, O*c* (Wolf-Rayet), O*d* and O*e* (characteristic of the hottest stars); below, the Sun for comparison, and types R and N (characteristic of the coolest stars).
(Photos, Curtiss, Ann Arbor Observatory of the University of Michigan)

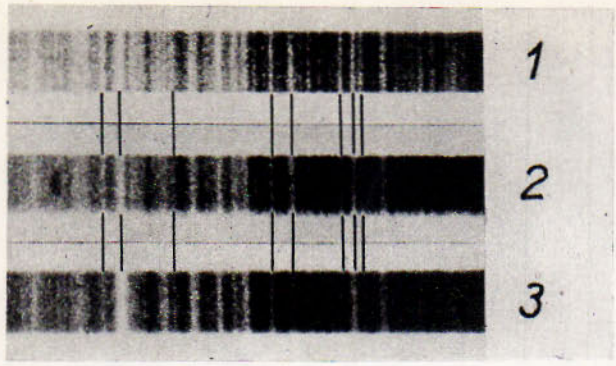


Fig. 610 — TYPICAL VARIATION OF THE RELATIVE INTENSITIES OF CERTAIN SPECTRAL LINES AS A FUNCTION OF ABSOLUTE MAGNITUDE. The three stars whose spectra (from λ 4200 to 4400) are here reproduced in negative are: 1, μ Cephei; 2, α Scorpii; 3, α Pegasi. They belong to the same spectroscopic type, M 1 — M 2, but the first two are supergiants ($M = -3.5$ and -3.0), while the third is an ordinary giant ($M = 0$). *(Keenan and Morgan)*

The G type stars raise an important distinction, which is applicable also to the preceding type and increasingly so to those that follow. On it is based the spectroscopic method of determining absolute magnitudes, to which reference has already been made in several different contexts. When the spectra of a giant and a dwarf of the same spectral type (Capella and the Sun, for instance) are compared carefully, it is found that the majority of the lines are finer in the giant's spectrum than in that of the dwarf, and also that the relative intensities of certain lines are different in the two spectra. By means of these differences (which we shall see later are a function of pressure) giants and dwarfs can immediately be distinguished from one another; they are denoted by the letters *g* and *d* respectively, prefixing the spectral type. Among the more notable G type giants is Capella (α Aurigae) *g*Go — in reality a spectroscopic binary, whose components are of types *g*G4 and *g*F4; and among the dwarfs, α Centauri A (*d*G4) and the Sun itself (*d*G2, according to the most recent determinations).

Giants far outnumber dwarfs in the star catalogues, but this is certainly a selective effect dependent on their greater intrinsic brightness; in reality, dwarfs are of much more frequent occurrence than giants, as can be deduced from the list of the nearest stars given on p. 310. This remark applies equally to types K and M.

Type K: Sometimes known as 'sunspot stars' on account of the resemblance between their spectra and those of the spots. Their temperatures range from about 5,000° down to 3,500°C (dwarfs 5,000° — 4,000°, giants 4,500° — 3,500°). Metallic lines, especially those of iron, are always intense in spectra of this type; the lines of ionised calcium attain their maximum at Ko-K2 and thereafter begin to fade; the hydrogen lines similarly continue to weaken; neutral calcium, on the other hand, becomes more prominent. In the later sub-types, molecular bands (especially of titanium oxide) begin to appear.

Notable examples of K type giants are Arcturus (α Boötis), *g*KO, and Aldebaran (α Tauri) *g*K5; and among the dwarfs, α Centauri B (*d*K1) and the two components of 61 Cygni (*d*K5 and *d*K6).

Type M: Titanium oxide stars. The most characteristic feature of these spectra is the fluted bands, fading on the red side, which are produced by titanium oxide, TiO. Temperatures from about 3,500° to 2,000° (dwarfs 3,500° — 2,500°, giants 3,000° — 2,000°C), which are insufficient to cause the dissociation of the more stable molecules. For the same reason, the normal lines of neutral metallic atoms are very numerous; those of iron and calcium are particularly intense. The lines of ionised calcium are still feebler than they were in the preceding type, while those of hydrogen are barely visible, except in certain stars (especially long-period variables — see Chapter 13) where they appear in emission.

Typical giants of this type are Betelgeuse (α Orionis) *g*M2, Antares (α Scorpii) *g*M1, and Mira (o Ceti) whose spectrum varies from *g*M5*e* to *g*M8*e*; the dwarfs are very numerous, and include Barnard's star (*d*M6) and the two components of Krüger 60 (*d*M4 and *d*M6).

Type S: The so-called zirconium oxide stars, from the abnormal abundance of this compound in their atmospheres. Apart from this feature, S type stars seem to be very similar to the giants of type M, and they may be regarded as constituting a rare variety of these. They are almost all variable, and practically none is visible to the naked eye, the two brightest being the long-period variables R Andromedae and R Cygni (magnitudes 6 — 14, period about 14 months).

Type R *and N*: The carbon stars, whose spectra are characterised by the bands of molecular carbon, C_2, and of cyanogen, CN, fluted on their violet side; their intensity increases steadily from Ro to R5 and from No to N3. The relative abundances of the various isotopes of carbon in these stars seem to be very different from the terrestrial: whereas carbon of atomic weight 12 predominates on Earth, carbon 13 appears in great quantities in the R and N stars. Their temperatures are difficult to determine, but probably range from about 3,000° to 2,000°C. They are all giants, and the majority are variable; they are of rare occurrence and none can be easily seen with the naked eye. Among the brightest are U Hydrae (N2), whose magnitude varies irregularly between 5 and 6, and the deep red 19 Piscium (No) of magnitude 6.

To this series of normal stellar spectra, two further types are sometimes added to look after certain peculiar categories of spectra: P, consisting of bright-line nebular spectra (oxygen and nitrogen several times ionised, and hydrogen), and Q, through which novae pass during the course of their rapid evolution, which usually takes them from B or A to W (Wolf-Rayet) via Q and P.

Having now learnt something about the luminosities, sizes and temperatures of the stars, we must next consider their masses: less technically, but also less accurately, what do the stars 'weigh'?

Masses and densities of the stars. — Remembering the enormous range of stellar luminosities and dimensions, we might expect to encounter a similar diversity among their masses: it would not be unreasonable to anticipate, for example, that these might be proportional to their volumes.

The only available method of determining stellar masses directly is through the study of the motions of the components of binary or multiple systems. We have already learnt that these motions conform with the law of universal gravitation and that accordingly we can deduce the masses of the bodies from the observations of their mutual motions, just as the mass of a planet can be accurately derived from the motions of its satellites.

The number of binaries whose orbital motions are sufficiently well known to allow of the calculation of their masses is still rather limited (less than a hundred in all), but fortunately almost all stellar types are represented, so that quite general conclusions can be drawn from them.

The surprising result that has emerged from such studies is that the amplitude of the variation of stellar masses is very restricted; whereas luminosities are spread

over more than a thousand million-fold range, and stellar volumes differ by an even larger factor (exceeding 10^{12}), the heaviest stars appear to be no more than 100 times as massive as the lightest. No dwarf is known to have a mass smaller than 0.1 times the Sun's, while stars with masses even 10 times the Sun's are supergiants of extremely rare occurrence.

If, now, the masses of the stars differ so little among themselves while their volumes are as varied as we have seen them to be, if follows that their densities (which are inversely proportional to their volumes for a given mass) must also vary over an extraordinarily wide range. To cite some examples: the Sun, with a diameter of 864,000 miles, has a mass of 2×10^{27} tons (or 2×10^{33} grams), corresponding to a mean density of 1.41 times that of water. In the case of a giant like Antares, on the other hand, whose diameter is about 480 times the Sun's, but whose mass is certainly no greater than 20 times that of the Sun, a quick calculation will show that its density is less than 2 ten-millionths that of the Sun, or roughly 0.0001 that of ordinary air.

The outer layers, at least, of so enormous and rarefied a globe must be to all intents and purposes completely transparent to the radiation passing through them from the lower levels of its interior, as well as that of any star happening to pass behind it. Such a state of affairs has actually been observed in the case of one of the largest stars known, the K4 supergiant ζ Aurigae: when, every 2½ years, its brilliant B8 *comes* passes into eclipse behind it, for several weeks at the beginning and end of the eclipse it can still be observed through the transparent outer layers of the supergiant, which only very gradually blot it out. The spectrum of the light that is thus filtered through the atmosphere of the primary contains the characteristic absorptions of these layers. This intriguing and important observation was made by O. C. Wilson in 1939, and for the first time permitted us to explore directly the atmosphere of a star — though admittedly under rather special circumstances.

Let us now consider the opposite extreme, as exemplified by a red dwarf smaller than the Sun: Krüger 60 B, for example, or one of the components of Barnard's star, whose mass and radius are both of the order of 0.2 times the Sun's. Its mean density will therefore be about 25 times that of the Sun, or 35 times that of water: this considerably exceeds that of platinum or iridium (with densities 21—22), the densest substances known on Earth. This, furthermore, is only its mean density — that at its centre will be even greater.

But there are still denser stars than this: the white dwarfs, whose prototype is the companion of Sirius. For 40 years or more, their existence has caused astronomers many a headache, and their nature is still a much discussed topic.

We saw in the last chapter that Sirius B was discovered, over a century ago, as a result of the perturbations that it causes in the proper motion of its primary. Since that time it has been studied assiduously, and its observed motions have allowed its mass to be derived with considerable accuracy: Sirius B is about as massive as the Sun. On the other hand its intrinsic luminosity is very low — certainly less than 0.01 times the Sun's and probably nearer to 0.003, though the proximity of the blindingly bright Sirius A makes accurate determinations difficult. Yet the companion's spectroscopic type (A5) indicates a surface temperature of 9,000° to 10,000°C. From these data it can be calculated that its diameter is only 0.02 that of the Sun, or twice that of the Earth. But for a body to have the mass of the Sun, and a diameter only twice that of the Earth, its matter must be compressed to an inconceivable degree: its density works out at more than 100,000 times the Sun's, or more than 150,000 times that of water.

WHITE DWARFS WHOSE ABSOLUTE MAGNITUDES WERE KNOWN IN 1941 (after G. P. Kuiper)

STAR	APPARENT VISUAL MAGNITUDE	ABSOLUTE MAGNITUDE (VISUAL)	SPECTROSCOPIC TYPE	COLOUR CLASS
Sirius B	8.5	11.4	A5	*a*
40 Eridani B .	9.6	11.1	A	*b*7
Procyon B ...	10.8	13.1	?	?
Wolf 1346 ...	11.3	9.8	A	*b*4
Hertzsprung 3	11.7	11.0	A	*b*5
— 7° 3632 A .	12.0	11.0	A	*b*5
van Maanen 2	12.4	14.3	G	*a*6
Luyten 745-6	13.1	14.3	F	*b*
AC 70° 8247..	13.6	12.3	Contin.	o
Hyades No. 16	14.0	11.2	A	*b*
Hyades No. 7	14.1	11.3	A	*b*
Ross 627	14.1	13.6	A5	*b*
Wolf 489	14.5	15.1	Contin.	*g*8
Wolf 219	15.2	14.3	Contin.	*b*
Wolf 457	15.6	15.1	Contin.	*a*

It is no unknown thing for Truth to appear in unlikely guises, and modern research has not only confirmed the reality of these unimaginable densities but has offered a theoretical explanation of them, though they still confound common sense. However, we must postpone the further discussion of this topic till the next chapter, which is devoted to the conditions prevailing in the interiors of the stars.

Until recent times it seemed that there was a gulf fixed between the self-luminous stars and the planets — non-luminous bodies shining faintly by borrowed light; Jupiter, the most massive planet in the Solar System, is still only about 0.001 as massive as the Sun, while white dwarfs, amongst the lightest known stars, have masses of about 0.1 the Sun's.

The discovery in recent years (mentioned in the last chapter) of planetary companions to certain double stars, whose masses are in the neighbourhood of 0.01 times the Sun's, has more or less bridged this gulf. It also raises the problem of the lower limit of stellar mass. We shall see later that there is a remarkable relationship involving stellar mass which allows us to fix this limit at several hundredths of the Sun's.

The Hertzsprung-Russell diagram. — So far we have been accumulating data relating to the luminosities, sizes, temperatures and masses of the stars, without bothering too much whether these are independent of one another or whether, on the contrary, they are related in such a manner that one determines all the others. It has, however, transpired that the giants are generally speaking the most massive and the same time the most luminous.

We must now discuss the relationships that link these various stellar characteristics one to another, enabling us to arrange the stars in sequence.

Half a century ago two astronomers, a Dane and an American, constructed the first of these, the diagram which is named after them being called the Hertzsprung-Russell diagram. In recent years this original diagram has been profoundly modified; indeed, it bears only slight resemblance to the latest suggestions. Nonetheless, it represents the foundation on which current theory has to a large extent built; we shall therefore consider Hertzsprung's and Russell's ideas in detail, since an appreciation of them is necessary for a proper understanding of the subsequent modifications.

Let us therefore begin by re-tracing the line of reasoning which led to the drafting of the Hertzsprung Russell diagram. Let us use the data in the tables on p. 310 to construct a graph whose ordinates (vertical axis) are absolute magnitudes — or what comes to the same thing, luminosities — and whose abscissae (horizontal axis) are temperatures, or, more conveniently, spectroscopic types. (We have already seen that the spectroscopic types are arranged in order of decreasing temperature.)

Thus a star plotted near the top of the diagram will be intrinsically very brilliant, while one at the foot will be intrinsically faint; one at the left will be very hot, and towards the right increasingly cool. The Sun (absolute magnitude + 5, temperature 6,000° C, type G2) will be situated roughly at the centre of the diagram.

On this basis, let us plot the positions of the stars listed on p. 310 We shall immediately see (Fig. 611) that, far from being scattered at random, they nearly all lie along or close to a line running obliquely across the diagram from the upper left to the lower right: that is, from type B stars with luminosities exceeding 100 times the Sun's towards low-temperature type M stars whose luminosities are less than one-hundredth of the Sun's. This line, along which luminosity and temperature decrease steadily and simultaneously, is known as the Main Sequence. The section of it that is occupied by stars less than 10 times as luminous as the Sun is known as the dwarf branch; the red dwarfs are to be found here, with luminosities less than 0.1 times the Sun's. We have already seen that these stars are extremely numerous, with a clear majority among the stars in space.

Several isolated stars in the lower left-hand corner of the diagram, well below the Main Sequence, will be noted. These are the celebrated white dwarfs, of which Sirius B and 40 Eridani were the first examples to be discovered. Although their faintness renders detection difficult, and less than a hundred are known, the search for more is being actively pursued, especially in America, and judging from the number already discovered in the neighbourhood of the Sun it seems that they are of frequent occurrence — probably more frequent than stars of the solar type. An investigation by the Russian astronomer P. Parenago (1946) suggested that the white dwarfs may be arranged in one or two series, roughly parallel with the Main Sequence (Fig. 613), for at the present time we only know of the yellowish members of the class, such as Wolf 489 (colour class *g*8).

The handful of stars scattered across the top of Fig. 611 are giants and supergiants of various spectroscopic types. Although their frequency, as displayed here, gives an exaggerated idea of their real spatial frequency (we have seen that they are in fact extremely rare) there are nevertheless too few to reveal the manner in which they are distributed on the Russell diagram. This, however, can be seen from Fig. 612, a schematic representation of the diagram which includes all classes of star. It will be observed that a fairly well-marked series of stars, with absolute magnitudes around zero (luminosities 100 times the Sun's), branches out horizontally from the Main Sequence at type A and extends as far as type M. This is the giant branch, which after type F (and still more obviously after type G) is clearly distinguishable from the dwarf branch. Here, in diagrammatic form, is the distinction that we have already referred to when describing the spectroscopic types, and which becomes more accentuated as we pass from type F towards type M. Thus a red type M star is either a high-density dwarf, several times smaller than the Sun and thousands or tens of thousands

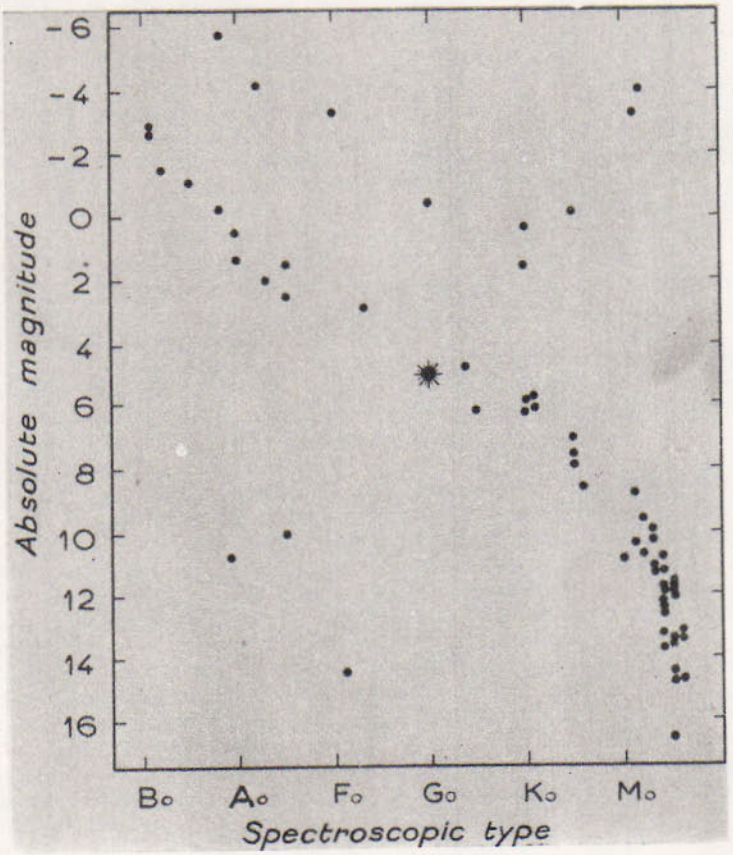


Fig. 611 — THE HERTZSPRUNG-RUSSELL DIAGRAM FOR THE BRIGHTEST AND NEAREST STARS. The Sun's position is marked by an asterisk near the centre of the diagram.

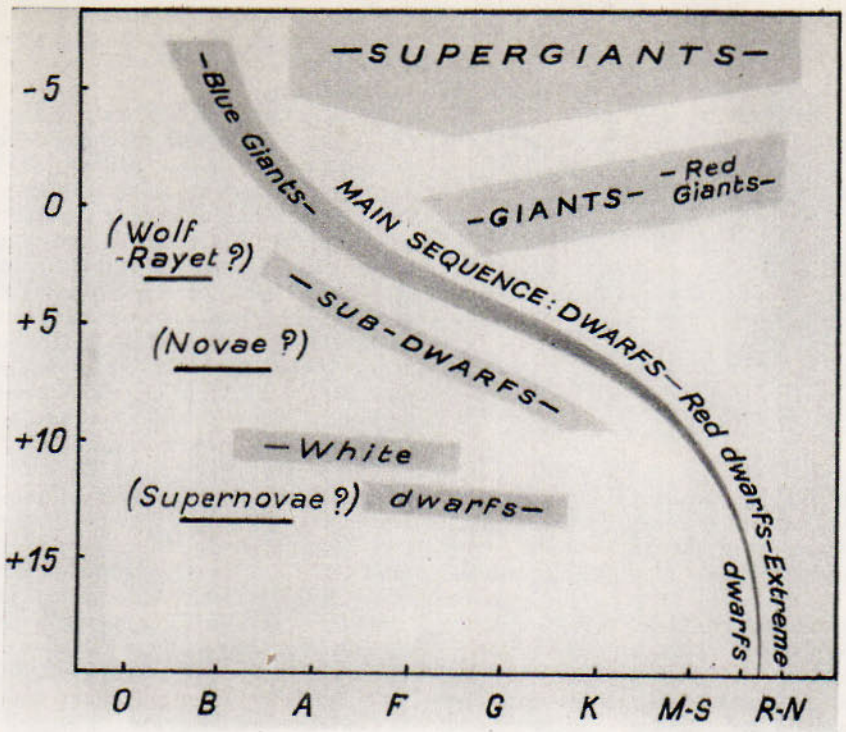


Fig. 612 — THE NAMES AND GENERAL DISTRIBUTION IN THE HERTZSPRUNG-RUSSELL DIAGRAM OF THE PRINCIPAL CATEGORIES OF STARS

of times less brilliant, or else a rarefied giant, ten or a hundred times larger, and also more luminous, than the Sun.

Finally a few supergiants, surpassing even the giants, are distributed rather irregularly across the top of the diagram. They are, however, slightly concentrated towards the top left-hand corner of the diagram (hot blue supergiants of types O and B) and the right-hand corner — cool red supergiants of types M-S and R-N, the majority, if not all of which, are irregular or long-period variables; they include some of the largest stars known, such as ζ Aurigae, Antares, Betelgeuse and Mira. Their enormous luminosities (several thousands of times greater than the Sun's) can only be explained, in view of their low intrinsic emission per unit area (due to their low temperature) by an equally gigantic radiating surface.

This leads to the question of how binaries and variables are distributed on the Hertzsprung-Russell diagram. Doubles, in general, may occur anywhere, and the most extraordinary associations of disparate stellar types are to be encountered. Typical examples of the association of similar stars, from various regions of the Main Sequence, are α Centauri, an association of two yellow dwarfs, types G4 and K1, comparable with the Sun; 61 Cygni, which consists of two orange-coloured dwarfs, K5 and K6, and Krüger 60, a pair of low-luminosity red dwarfs, M4 and M6. More ill-assorted pairs include Sirius, a system consisting of a normal Ao star, larger and more luminous than the Sun, and a minute A5 or A7 white dwarf, 10^4 times less luminous than its primary, but of practically equal mass; o_2 Eridani, which is a bizarre association of a normal G5 star, a B9 white dwarf (6 times less luminous), and an M5*e* red dwarf (300 times less luminous); and VV Cephei, a cool supergiant of low density and temperature (1,500°), with a very hot and luminous B type companion.

In general, however, there is a tendency for visual binaries to occur most frequently among the Main Sequence dwarfs, and for spectroscopic binaries (including eclipsing variables) to show up among the Main Sequence giants and supergiants, though there are many

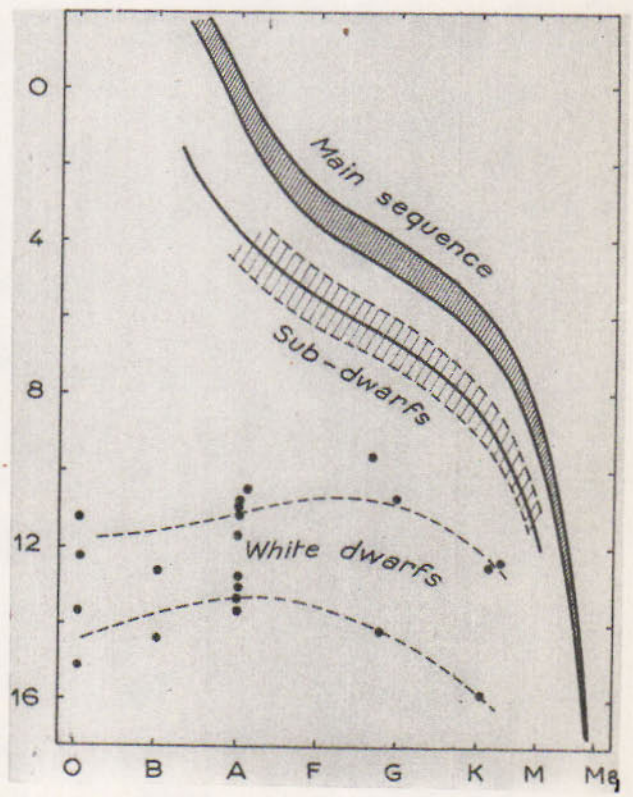


Fig. 613 — POSITION OF THE SUB-DWARFS AND THE TWO PROBABLE SERIES OF WHITE DWARFS IN THE HERTZSPRUNG-RUSSELL DIAGRAM. *(P. P. Parengo, 1946)*

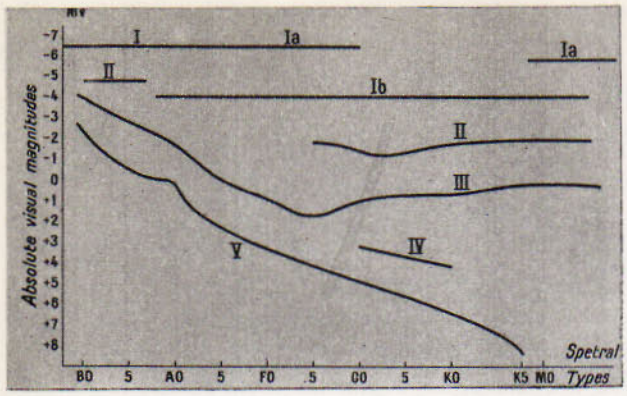


Fig. 614 — MORGAN'S DIAGRAM, WHICH TODAY TAKES THE PLACE OF THE HERTZSPRUNG-RUSSELL DIAGRAM

exceptions to this rule; furthermore, the W Ursae Majoris eclipsing variables appear to be mostly dwarfs, though this preference is probably more apparent than real, being imposed by selective factors connected with observational technique.

The distribution of variables, so-called, is more specific. Cepheids are all giants of great intrinsic brightness; cluster variables (those with periods shorter than 1.5d) are white B, A or F giants with absolute magnitudes around 0; classical Cepheids are yellow giants or supergiants of types F, G or K, whose absolute magnitudes increase with their periods at the same time as their spectroscopic type progresses; and the long period Cepheids (periods exceeding 100d) are extremely luminous K or M supergiants. Only the β Canis Majoris Cepheids, or dwarf Cepheids, belong to the Main Sequence (types B, A of F).

Long-period variables, semi-regular and irregular variables are all 'late type' red giants or supergiants of types K, M, R, N and S. Among the peculiar variables some, e. g. R Coronae and R Aquarii, are likewise yellow or red giants or supergiants; others, e. g. γ Cassiopeiae, P Cygni, and the stars of the Pleiades, are very hot type B giants or supergiants.

Novae, after their outburst, appear to be hot O or W (Wolf-Rayet) stars, and perhaps B, A or F stars before it, though too few spectra of novae before maximum have been obtained for us to be very sure of this. But since their absolute magnitudes at maximum are in the region of -6 or -7, and their mean amplitude 11 — 12 magnitudes, their normal absolute magnitude of $+5$ placed them at first in an apparently otherwise unoccupied region of the Hertzsprung-Russell diagram. Later, however, certain other stars were discovered whose position was a little below the Main Sequence dwarfs, making them look like an intermediate class between normal dwarfs and white dwarfs. It appears that novae are recruited from the ranks of these subdwarfs.

It is difficult to say much about the supernovae owing to our virtual ignorance of their amplitudes. If they are blue stars (as SN 1054 and SN 1572 would suggest) and their amplitudes are of the order of 20 magnitudes, their pre-outburst absolute magnitudes must be in the region of $+5$ or more, which again would put them among the sub-dwarfs, if not among the white dwarfs. But more will be said regarding modern theories of the formation of novae and supernovae in Chapter 18.

The Morgan diagram. — As will be readily realised, the original Hertzsprung-Russell diagram, while sufficing to demonstrate the existence of different luminosity classes, suffered from a number of serious drawbacks. Only stars whose distances (or what amounts to the same thing, absolute magnitudes) are known can be plotted on it, which implies that the narrowness of its plotted zones will therefore be directly determined by the accuracy of our stellar distance determinations, and this is steadily improving. There were no narrow zones in these early diagrams, but only 'clouds' of points; the median axes

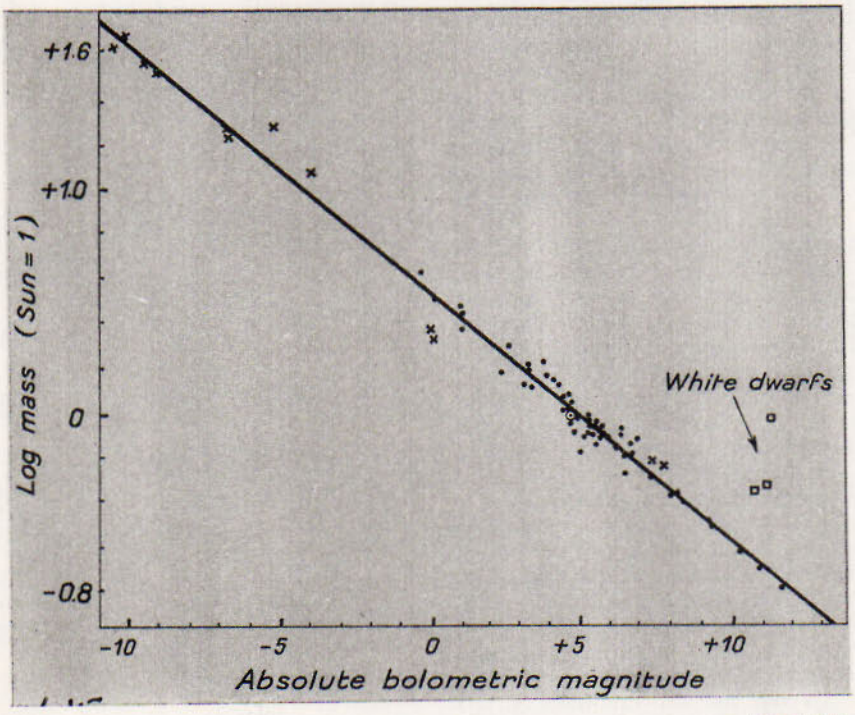


Fig. 615 — EMPIRICAL RELATIONSHIP BETWEEN THE MASSES AND THE ABSOLUTE MAGNITUDES OF THE STARS. Visual binaries; x Spectroscopic binaries; ⊙ Sun; □ White dwarfs. *(P. Baize, 1945)*

of these coincided with the zones as nowadays recognised, but their spread on either side amounted to one or two absolute magnitudes.

A recent series of investigations undertaken at Yerkes Observatory (in particular, those under the direction of W. W. Morgan) has refined these data very considerably.

Morgan's diagram (Fig. 614) recognises five different classes of stars; the new ones, II and IV, are respectively intermediate between supergiants and giants, and between giants and dwarfs. It should be mentioned that the curves corresponding to these classes are not more than about 0.5 absolute magnitudes wide, which represents a very notable gain in precision.

A simple calculation will show that this degree of uncertainty in the absolute magnitude of a star corresponds to an uncertainty of 10% in its distance. The widespread use of photoelectric techniques for the measurement of apparent magnitudes encourages us to anticipate a reduction of this margin of uncertainty to ± 0.05 absolute magnitudes, or ±1% in the corresponding distances.

The following elementary facts concerning the Russell and Morgan diagrams should be borne in mind. An astronomer who is experienced in such work has no difficulty in identifying the type of a star from its spectrum. But to each such spectroscopic type correspond several possible values of the absolute magnitude; criteria must therefore be established which will permit the luminosity class of a star to be determined. Such criteria are based on the following considerations.

(1) Since ionised atoms are more abundant in a giant or supergiant than in a dwarf, the corresponding lines will be enhanced in the spectra of the former, as compared with those of neutral atoms of the same elements.

(2) If one examines the spectrum of a given element in several different stars, it will be found that its absorption lines are much finer in a supergiant than in a dwarf; the fineness of the lines is in this case directly related to the very weak pressures obtaining in the outer layers of the former.

These various considerations boil down to the following fact: the examination of the spectrum of a star allows both its spectroscopic type and its luminosity class to be determined (with an accuracy which is being constantly improved), and from the Morgan diagram its absolute magnitude, and hence its distance, can then be derived.

These are the essentials of the method whereby stellar distances are measured by purely spectroscopic techniques, with an accuracy of the order indicated above.

Until recently this method, which has made enormous progress over the last twenty years, was based on the establishment of criteria corresponding to the relative intensities of certain standard pairs of lines (one in the arc spectrum and the other in the spark). A calibration curve for each type was then constructed. (This method is described more fully on p. 496).

This brief survey would be incomplete if the following points were omitted.

(1) The scope of the method is limited by the necessity of obtaining spectra whose dispersion and clarity are both adequate.

(2) It can be seen from Fig. 614 that Morgan's diagram does not extend to stars of type M and cooler. The classification of the coolest stars involves problems whose satisfactory solution is not yet in sight.

Recent work has also shown that the Russell and Morgan diagrams are inapplicable to certain other categories of star — among them, in particular, the members of globular clusters and above all the short-period RR Lyrae Cepheids which we have already discussed. Furthermore these stars possess other distinguishing characteristics, and in fact belong to a distinct stellar population (Population II). This is a very recent discovery; it has had a great influence on our theories of the evolution of the Universe, and is best treated in that context. Accordingly, we will defer a discussion of it until we come to the study of cosmic evolution (Chapter 18), when we will deal with the modified diagram which replaces the Hertzsprung-Russell one for these stars.

As we have mentioned above, for each spectroscopic type there are in general several possible values of the absolute magnitude. Some astronomers have been led to suspect that this division into separate luminosity classes

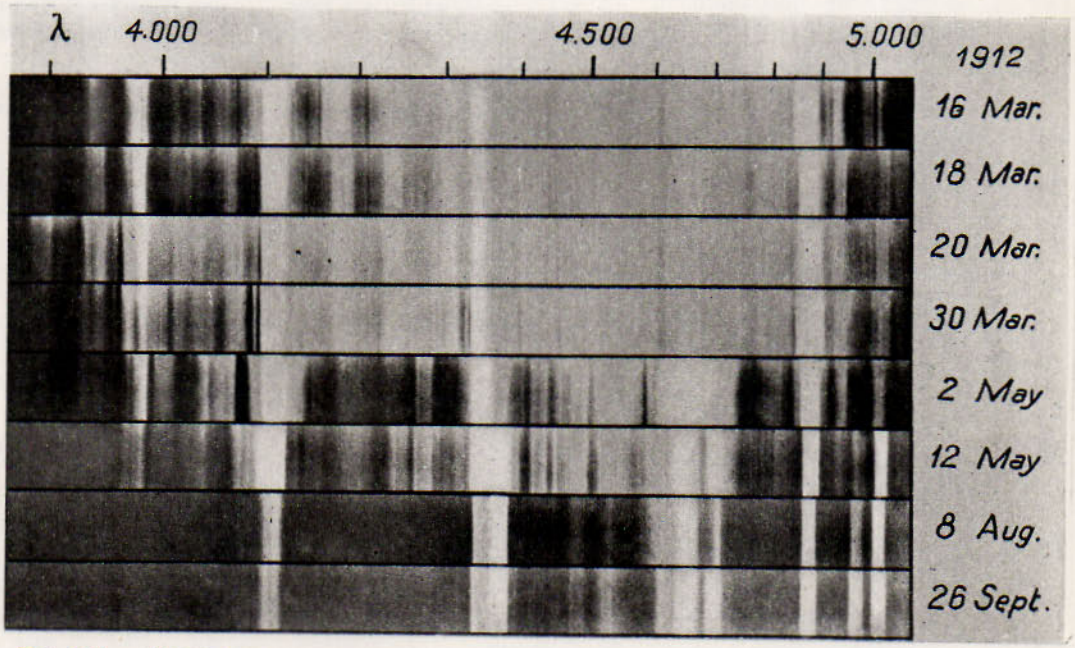


Fig. 616 — THE SPECTROSCOPIC CHANGES OF NOVAE (EXEMPLIFIED HERE BY NOVA GEMINORUM 1912) TELL THE STORY OF THEIR EVENTFUL LIFE. They form a characteristic series which, whether it is passed quickly or slowly, always follows the same sequence — the evolution of a type B or A spectrum into a bright-line gaseous spectrum. *(Photo, Lick Observatory)*

is really nothing more than a selective effect imposed by our methods of observation, and that the introduction of new techniques will eventually reveal stars linking any given absolute magnitude with each spectroscopic type. The rectangle of the Russell diagram would then be completely filled with points, each representing a single star. But only the future can tell whether this hypothesis is well founded.

The mass-luminosity relation. — The Hertzsprung-Russell diagram introduced a considerable degree of order into the diversity of stellar types; we have now seen that the development of it which we owe to Morgan's work has still further improved this. Nonetheless, the situation remains sufficiently complicated, what with giants and supergiants, dwarfs, sub-dwarfs and white dwarfs, and though this system of classification gives hints, it gives no definite information regarding any fundamental relationships which may link these diverse categories of star. Such a relationship has, however, been discovered by comparing the luminosities of the stars not with their temperatures but with their masses (as derived from the study of binaries).

Up to this point, when speaking of luminosities and absolute magnitudes we have been referring to the visual rating, but in discussing the relation between mass and luminosity we shall be dealing with bolometric absolute magnitudes — that is, visual absolute magnitudes to which a correction has been added to take into account the ultra-violet and infra-red radiations to which our eyes are not sensitive, and also those radiations that are absorbed in the Earth's atmosphere. This correction amounts to more than −1 magnitude in the case of red stars.

With this qualification in mind, we can see from Fig. 615 (based on the best observational data, and collected by the French amateur astronomer P. Baize in 1945) that there is a very close connection between the masses and the luminosities of the stars — the more massive a star, the greater its luminosity — at any rate so long as only Main Sequence stars are considered. This mass-luminosity relation, discovered by Eddington in 1924, reflects a deep-seated property of the stars, and we shall have more to say about it, and particularly about its theoretical explanation, in the next chapter.

DOUBLE STARS ILLUSTRATING THE EMPIRICAL MASS-LUMINOSITY RELATION (after P. Baize, 1945)

STAR	SPECTROSCOPIC TYPE	MASS (⊙ = 1)	BOLOMETRIC ABSOLUTE MAGNITUDE	LUMINOSITY (⊙ = 1)
Y Cygni A, B	O9	17.2	− 6.6	36,000
μ Scorpii A, B	B3	12.4	− 3.8	2,750
Capella A	*g*G4	4.2	− 0.5	130
Sirius A	A0	2.4	+ 1.1	30
Sun	*d*G2	1.0	+ 4.8	1
61 Cygni A	*d*K5	0.7	+ 7.4	0.09
Krüger 60 A	*d*M4	0.24	+ 10.1	0.008
Krüger 60 B	*d*M6	0.16	+ 12.3	0.001
Wolf 359	*d*M8	(0.09)	(+ 14.2)	0.0002

The preceding table gives some concrete examples of the mass-luminosity relation.

The data of Wolf 359, which is not a double star, were obtained by extrapolation by means of the mass-luminosity relation, and are inserted here to show that even so faint a star as this — its total radiation is only 0.0002 that of the Sun — is still about one-tenth as massive as the Sun. It thus seems, as we have said above, that the faintest bodies that can be called stars nevertheless have masses of the order of several hundredths that of the Sun, and are accordingly considerably more massive than the recently discovered 'planetary' *comites*; it is quite possible, however, that future discoveries will close this gap.

Although our knowledge of the faintest stars is not yet complete enough to fix the lower limit of stellar mass with certainty, its upper limit is irrevocably determined by the corresponding limit of luminosity. We know that, at this end of the luminosity range, we are on fairly safe ground in stating that no star could have escaped notice if its bolometric luminosity exceeded the Sun's by a factor greater than about 10^6. In other words, as the mass-luminosity relation shows, no stable star has a mass much exceeding 50 times the Sun's. In the next chapter we shall discover Eddington's very simple explanation of this remarkable fact.

Thus despite the enormous diversity of stellar luminosities and volumes, their masses are all more or less comparable with the Sun's: certainly none is more than 50 times heavier than the Sun, and probably very few are more than 50 times lighter. The vast majority of stellar masses are restricted to an even narrower range than this, from about 0.2 to 5 times the solar mass.

Finally, it will be seen from Fig. 615 that three stars depart conspicuously from the mass-luminosity relation to which all the other stars adhere so obediently. These stars, which are relatively too faint for their masses, are Sirius B, Procyon B and 40 Eridani B — white dwarfs, which thus manifest their fundamental structural dissimilarity from other stars.

The existence of such a relationship as that between mass and luminosity, and the almost universal scope of its validity (from blue giants to red dwarfs, with only the white dwarfs abstaining) have played a leading part in the last quarter century's theoretical investigations of the internal constitution of the stars. To this subject we must now turn.

Fig. 617 — THE STARS, LIKE THE SUN, ARE GIGANTIC FURNACES MAINTAINED BY ATOMIC ENERGY

CHAPTER 15

THE NATURE OF THE STARS

In the last few chapters we have become acquainted with those 'public' properties of the stars which are given by observation — temperatures, masses and luminosities; behind their apparent diversity lies hidden the fundamental homogeneity whose expression is the mass-luminosity relation. It is now time to leave the purely observational field and to seek an understanding of the deep-seated structure of the stars, as it is reflected in their masses, compositions and evolutionary stages.

This subject, involving the application of the fundamental laws of modern atomic and molecular physics, can be dealt with only superficially in a general work such as the present, but it is hoped that enough will be said to give some idea of the progress that has been made during the past thirty or forty years — the effective span of the study of stellar structure.

THE CONSTITUTION OF THE STARS

Stellar atmospheres. — In Chapter 10 we described the brilliant surface of the Sun — the photosphere — and its overlying and increasingly rarefied atmospheric layers — the reversing layer, the chromosphere and the corona; and it has been implicitly assumed, when speaking of the stars, that their atmospheres are organised according to a similar scheme. It is now necessary to enquire more precisely what we mean by such terms as 'surface', which have by no means the same significance as when, for instance, they are applied to the Earth. In the latter case the distinction between the solid or liquid surface and the overlying gaseous atmosphere is perfectly clear and unambiguous. The same is not true of the Sun and the stars.

We must regard these bodies as wholly gaseous spheres, since their high temperatures prohibit any element from remaining in the solid or liquid state.* Towards the centre of a star the gases will be highly compressed by the weight of the overlying layers, as will be explained in a moment; their density may reach values far exceeding those of the densest metals. The radiant energy generated in these central regions, while making its way towards the star's surface, is absorbed at intermediate levels and is thus prevented from being propagated directly to the exterior; the density and temperature of the star's matter decrease in proportion as its distance from the centre increases, since the thickness and therefore the weight of the down-pressing overlying layers decreases steadily; a point is finally reached at which the radiation is able to filter out towards the star's exterior. The radiation from the outermost layers of the star, on the other hand, can emerge without difficulty. At the same time, the radiation of these outer layers is weaker in proportion as their temperature is lower; it also differs from the radiation of the deeper and hotter levels in its wavelength distribution, being relatively richer in red light. Finally, the outermost layers of all neither emit nor absorb any great quantity of radiation, and in these regions the increasingly rarefied matter of the star's atmosphere tails off imperceptibly into the interstellar void.

Under such conditions as these it is clearly misleading to talk about a well defined 'surface' as producing the light which we receive on Earth: rather, this light is emitted from a zone of considerable thickness, whose depth depends on the opacity of its constituent gases — depends, that is, on their temperature and pressure rather than on their chemical composition, which is a factor of only secondary importance.

Thus it is impossible to draw a clear-cut distinction

* Save only carbon, in the coolest stars: see Chapter 13, long-period variables.

between a star's photosphere and, on one hand, its interior, or, on the other, its increasingly rarefied atmosphere. The photosphere can only be defined as the totality of the layers emitting radiation which is directly observable. Its thickness depends essentially, for a given spectroscopic type, on the force of gravity at the star's surface. The density and opacity of a highly compressed white dwarf increase very rapidly with increasing depth; consequently the layer emitting observable radiation will not be very thick — a few hundreds or thousands of miles at the most: it will, in fact, have a comparatively sharply defined photosphere. But the gases of a highly rarefied giant retain their transparency to great depths — some 10^5 or 10^4 miles — and the photosphere cannot be said to have any definite boundaries.

By way of example, here are data of the atmospheres of two G type stars, as derived by the French astronomer D. Barbier: one is a dwarf (the Sun) and the other a giant such as Capella; the effective observed temperatures of both is 5,740° K (i.e. absolute temperature). The force of gravity at the surface of the dwarf is 28 times that at the Earth's surface; towards the base of the chromosphere (or more precisely, in the layer producing the star's integrated radiation) the true temperature is 4,880°, the pressure exerted by free electrons is 1.4×10^{-6} atmospheres, and the total pressure of the gas is 0.01 atmospheres, corresponding to a density of 2×10^{-5} that of the air at sea level. But only one-third of the radiation emitted by the layer 180 miles below this reaches the exterior; in this layer the temperature has risen to 5,600°, the electron pressure to 1.5×10^{-5} atmospheres, and the total pressure to 0.085 atmospheres, with a corresponding density 1.4×10^{-4} that of our atmosphere at sea level. Finally, at a depth of about 370 miles the temperature rises to 7,200°, the electron pressure to 2.8×10^{-4} atmospheres, the total pressure to 0.15 atmospheres, and the density to 2×10^{-4} that of air; but only 0.2% of the radiation from this level ever manages to make its way to the surface of the star and launch out into space. It may therefore be regarded as the effective lower limit of the photosphere, which thus has a total thickness of something under 400 miles.

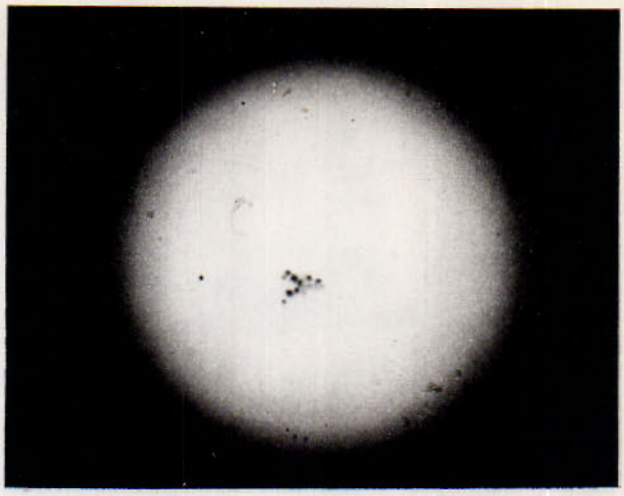


Fig. 618 — This photograph of the Sun, taken in blue-violet light, clearly shows the marked difference in brightness between the limb and the centre of the disc. *(Photo, L. Rudaux)*

In the case of the giant, whose surface gravity is no greater than that of the Earth, the thickness, of the effective photosphere is 20 times as great, or about 7,000 miles, and its density and pressure at equivalent depths are of course correspondingly smaller.

This picture provides a simple explanation of the phenomenon which has already been described in Chapter 10

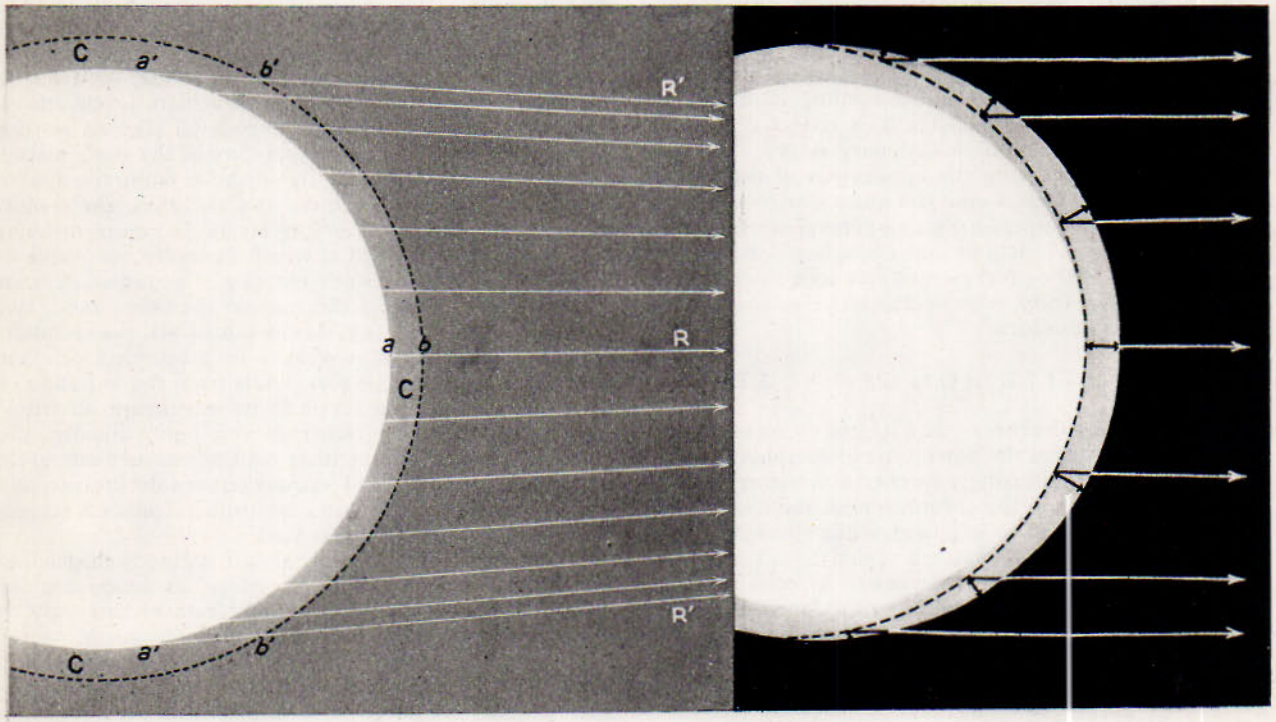


Fig. 619 — It was originally thought that the darkening of the Sun's limb was due to the greater thickness of atmosphere traversed by the rays RR′ from the limb *a′b′*, than that at the centre, *ab* (left). It is now understood that the limb-darkening effect springs from the varying effective level of the photosphere (right), whose optical depth is everywhere the same, and which is therefore deeper in the Sun (and hence hotter) at the centre of the disc than at the limb.

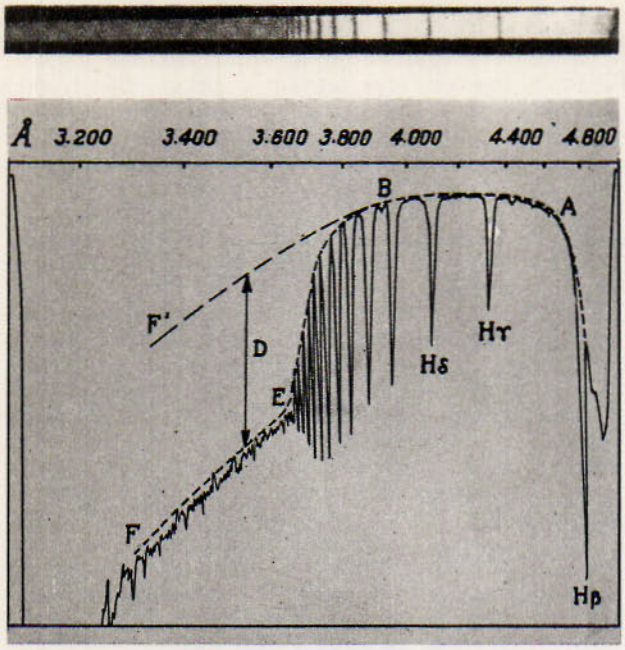


Fig. 620 — This spectrum of Vega (α Lyrae), above, and its microphotometric trace, below, show clearly the importance of the hydrogen absorptions in the ultra-violet regions of type Ao stars.
The sharp points on the curve correspond to the absorption lines Hβ, Hγ, Hδ ... of the Balmer series of hydrogen. The distance D between the prolongation (B—F') of the spectrum's continuous background in the blue and violet regions (A—B) and the actual continuous background in the ultra-violet (E—F) is known as the Balmer gap. *(D. Barbier and D. Chalonge, Paris Observatory)*

— the darkening of the Sun's disc towards the limb, which is much more noticeable photographically than visually, i.e. is more marked in violet and ultra-violet light than in yellow or red. Fig. 619 shows quite clearly that the light which we receive from the limb regions comes from a more superficial layer than that from the centre of the disc, simply because light rays emitted from deep levels can only reach the observer if their paths through the overlying layers do not exceed a certain limiting value — the distance to which we can penetrate the Sun visually when looking at the centre of the disc. And since the light from the limb regions is generated in less deep and therefore cooler layers, it will be at the same time less intense and redder — which is exactly what is observed. Thus it is not because the photosphere is overlaid by an absorbing atmosphere that the centre of the disc is brighter than the limb, but because the mean level (and hence the temperature) of the effective photosphere within the atmosphere, which both emits and absorbs radiation, varies across the disc from the centre to the limb.

It will be appreciated from the foregoing account that the measurement of the limb darkening in different wavelengths will yield information about the opacity of the solar or stellar atmosphere to different radiations, and this is a most valuable datum, since it determines the star's emission of light to the exterior. The limb darkening of the Sun has been known and measured for a long time, but it is only in comparatively recent times that it has been detected among the stars, by means of the very detailed study of the light curves of eclipsing binaries during eclipses; it is obvious that if the disc of the eclipsed star is darkened towards the limb, the light variation will be more gradual than if it were uniformly bright.

The intrinsic opacity of stellar atmospheres seems to be very considerable — far greater than that of the Earth's atmosphere in clear weather, and more nearly resembling that of a thick fog. On the Sun, for example, a column 1 square centimetre in section, running through the 400-odd mile thickness of the photosphere, would contain only 3 gm. of matter, yet it would absorb 99.8% of the radiation from the underlying layers; a similar column of the Earth's atmosphere contains 1,033 gm. of air but normally absorbs no more than 20% to 30% of the radiation passing through it.

This very high opacity of the atmospheres of the stars is due primarily to the intense absorption which accompanies the excitation or ionisation of atoms. All wavelengths shorter than a certain critical value (which varies from element to element) are capable of being absorbed by a given atom, producing a rearrangement of its electronic pattern (excitation) or even removing one or more electrons from it altogether (ionisation). Thus neutral hydrogen is responsible for very strong absorptions in the ultra-violet regions of the spectra of types B, A and F; an analogous feature of the visible and infra-red regions of A, F, G and K type spectra (and particularly of the solar spectrum) has been detected by R. Wildt in America and investigated by D. Chalonge in France; it involves the negative ions of hydrogen — i.e. hydrogen atoms which have captured an extra electron* — whose dissociation (one can hardly call it ionisation, since the atom is returning to its neutral state) causes intense absorption of the light promoting it.

Additional causes of the opacity of stellar atmospheres are the ionisation of other elements, particularly metals, and the wide and characteristic absorptions due to molecules (titanium oxide, carbon, and various hydrides and oxides) in the K, M and later types.

It is important to realise that this chemical diversity of stellar atmospheres, revealed also in the descriptions of the different spectroscopic types in Chapter 14, is more apparent than real: it results primarily from temperature differences, which determine the radiations that the various atoms can emit at their different energy levels (see Chapter 20). Analysis of the intensities of the spectral lines (allowance being made for the effects of temperature) has shown that the atmospheres of the majority of the stars have virtually the same composition as that of the Sun's — which is naturally the best known. The following table gives the relative abundances of the principal constituents of the solar atmosphere, by volume and by weight, according to D. H. Menzel of Harvard College Observatory.

ELEMENT	VOLUME (%)	MASS (%)
Hydrogen	81.76	53.9
Helium	18.17	44.9
Oxygen	0.03	0.45
Magnesium	0.02	0.45
Silicon	0.006	0.13
Nitrogen	0.01	0.09
Sulphur	0.003	0.05
Carbon	0.003	0.02
Iron	0.0008	0.03

The preponderance of hydrogen and helium — together comprising 99.9% of the Sun by volume, or 98.8% by weight — is overwhelming. Appreciable quantities of

* The neutral hydrogen atom has a single electron.

the common terrestrial elements oxygen, magnesium, silicon, nitrogen, sulphur, carbon and iron also occur on the Sun; other elements are only negligibly represented. Since much the most intense lines of the solar spectrum are those of ionised calcium, and the most numerous those of iron, it will be realised that the true abundance of a given element cannot be deduced directly from the number or the intensity of its lines; the rôle of temperature, favouring or inhibiting its emission of radiation, is all-important. Nevertheless, it does seem that the atmospheres of certain very hot stars (type O Wolf-Rayets) and very cool ones (types M, R, N, S) are abnormally rich in certain elements such as carbon, strontium, silicon or sulphur. But these anomalies only concern elements of secondary importance, and it is essentially true to say that the stars are vast globes of hydrogen and helium. We shall see later that this preponderance of the two lightest and structurally simplest elements is a general characteristic of the stellar universe, and one of the greatest cosmological significance.

Stellar interiors. – Other branches of astronomy had advanced a long way before the mystery of the internal constitution of the stars began to be probed, for although it is comparatively easy to obtain information about their luminous atmospheres by spectroscopic analysis, it is obviously more difficult to learn anything of their interiors, from which we receive no direct radiation whatever.

Despite this initial difficulty, great progress of our understanding of the internal structure of the stars has been made by the theoreticians during the past forty years or so, through the intelligent application of general laws of matter and radiation whose universal validity has been firmly established by the physicists.

The first steps were taken by the American, H. Lane, as long ago as 1870, and further progress was made by the Germans Schwarzschild and Emden around 1906; but it is above all to Eddington, between 1916 and 1926, that we owe the real development of the theoretical investigation of stellar interiors and its consolidation into a single coherent theory. In recent years Eddington's work has been modified in various details by other workers – notably Chandrasekhar, an Indian physicist working in America.

Observation proves that the stars are gaseous spheres, the temperature and density of whose atmospheres rise rapidly with increasing depth. A simple argument shows that this increase continues until the centre of the star is reached. The matter at any given point in a star's interior must support the weight of the column of gas rising above it to the star's surface; this weight increases rapidly with increasing depth. Now the only two forces that are available to resist the collapse of the gaseous medium under its own weight are gas pressure (produced by the thermal agitation of its atoms) such as supports the Earth's atmosphere, and radiation pressure (due to the radiation which in a star tends to be propagated from the interior towards the exterior). Both these pressures rise sharply with temperature, but since the pressure exerted by the overlying layers continues to increase steadily from the surface to the centre of the star, the two opposing forces must likewise increase if the star is not to collapse under its own weight. In other words, the temperature must also increase steadily as the centre of the star is approached. We have already seen that the highest temperatures are to be encountered in the central regions of the stars, where the greater part of their energy is produced; the mechanism of this production will be described in detail later.

The energy so produced is propagated little by little towards the star's exterior, keeping the temperature of the intermediate layers at a value necessary to maintain them in stable equilibrium; for this reason the stars are said to be in a state of radiative equilibrium, since it is the absorption of radiation from the central region by the less deep layers that maintains them at the equilibrium temperature. To say that the radiation is absorbed by the intermediate layers is equivalent to saying that they are opaque to it; this is, indeed, fortunate, for were it otherwise the stars' energy resources would be rapidly depleted, and they would burn themselves out in a brief but violent holocaust.

The enormous quantity of energy that the stars radiate into space (we shall see shortly what its value is) is thus no more than a tiny fraction of the total energy produced in their interiors – the fraction that has succeeded in filtering through to the surface after countless successive absorptions and re-emissions. All the rest is absorbed in the interior, and serves to support the outer layers of the star – directly by radiation pressure and indirectly by gas pressure.

The calculation of this state of equilibrium, by means of the application of the gas and radiation laws, reveals that the central temperature of the stars must be several millions or tens of millions of degrees. At such temperatures the thermal agitation of the atoms would reach velocities of the order of some tens or hundreds of m.p.s.; the radiation emitted in these circumstances would consist of X- and γ-rays, whose photons are endowed with very high energies.

Laboratory work has proved that when atoms are bombarded with very intense radiation and constantly suffer violent collisions with their neighbours, they will lose most of their orbital electrons – i.e. they will be ionised, more or less completely (under given conditions) according to their individual powers of cohesion, but in any case to an advanced degree under stellar conditions. In the interiors of stars they are for the most part reduced to their nuclei, or at best retain only one or two electrons. At the centre of the Sun, for example, where the temperature is about 2×10^7 degrees, the average iron atom has lost all but 2 of its normal complement of 28 electrons. In this state an atom occupies much less space than the neutral atom – just as the Solar System would shrink

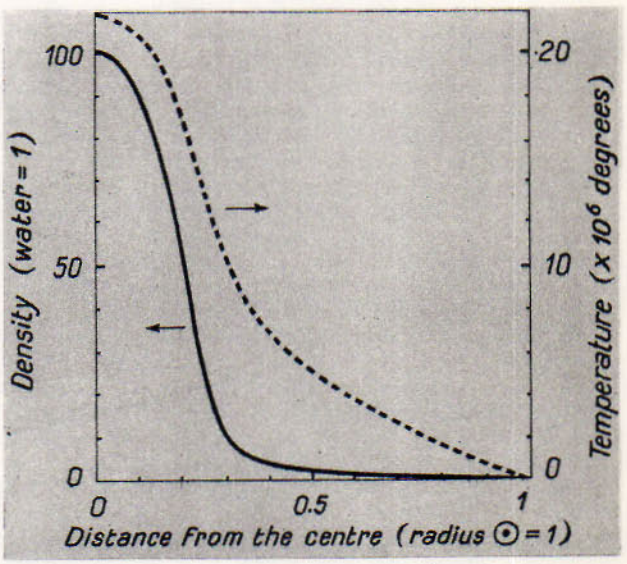


Fig. 621 – VARIATIONS OF TEMPERATURE AND DENSITY IN THE INTERIOR OF THE SUN. Note the high degree of central condensation.

in size if it lost all its planets and was reduced simply to the Sun. It is because of the highly ionised state of stellar material that the gas laws — and even those relating to a perfect (i.e. highly rarefied) gas — can be applied to stars like the Sun, despite the fact that their mean densities exceed that of water and their central densities that of the heaviest metals. At one time this application of the gas laws to the stars seemed not only inadmissible but absurd. It is because atomic nuclei and free electrons are of the order of 10,000 times smaller than the neutral atoms with their full complement of orbital electrons, that a given quantity of highly ionised material occupies something like a millionth of the volume that it would in the normal state. It thus constitutes a medium which, although it is 100 or more times as dense as water, retains the properties of a rarefied gas in which the individual particles can move freely and without mutual interference.

The ionised condition of the internal matter of a star has another important consequence. The pressure of a gas is in fact proportional to the number of particles per unit volume, as well as to its temperature. Now owing to ionisation, each single atom is replaced by a nucleus and a number of free electrons — 1 in the case of the hydrogen atom, 2 for helium, 6 for carbon, etc. In other words, one original particle is now replaced by 2 (hydrogen), 3 (helium), 7 (carbon), etc., and all of these contribute to the pressure; hence a given pressure is attained at a much lower temperature than would be the case if the atoms were not ionised. Furthermore — and this is a very important point — it follows that it is not necessary to know the exact chemical composition of a star's interior: the pressure of a gas at a given temperature is a function of its molecular weight, and since the number of electrons in an atom is very nearly equal to half its atomic weight* the mean 'molecular' weight of the totally ionised atom is always very near to 2, no matter what element it may be. The only exceptions to this are hydrogen and helium, for which it is 0.5 and 1.3 respectively, the mass of the atom being distributed among $n+1$ derivative particles. Thus for carbon $\frac{12}{6+1} \simeq 2$, for nitrogen $\frac{14}{7+1} \simeq 2$. for iron $\frac{56}{28+1} \simeq 2$, etc, while for hydrogen $\frac{1}{1+1} = 0.5$ and for helium $\frac{4}{2+1} = 1.3$.

For this reason the central temperatures and densities of the stars are primarily determined by the relative abundances of hydrogen and helium. The central temperature of the Sun would be 10 million degrees if it were composed solely of hydrogen, about 27 million if composed solely of helium, and 40 million if composed of any other element.

The actual percentage occurrence of these various elements on the Sun is derived by comparing the calculated energy output in each case with the observed output. In this way it has been established that the Sun consists of approximately 50% hydrogen and 40% helium, with other elements contributing 10% or less. The corresponding central temperature of the Sun is between 20 and 25 million degrees C, and its pressure about 10^{10} atmospheres, giving the gas a density of about 100 times that of water. In the same way it has been found that the central temperatures of the other Main Sequence stars range from about 10 million degrees (type M red dwarfs) to 30 or 35 million (type O and B giants); these enormous temperatures are imposed by the opacity of the stellar material. Along the giant branch from G to M, on the other hand, the temperatures are markedly lower, varying from a few million to a few hundred thousand degrees C only, since the incomparably lower densities of these stars prevents the retention of their energy to the degree achieved by the dwarfs.

One particularly remarkable outcome of these investigations into the internal state of the stars is the theoretical explanation (with a minimum of assumptions) of the mass-luminosity relation. This relation was first established by Eddington in 1924, the stars being assumed to consist of perfect gas; later it was shown to apply equally to the much denser dwarfs. This unexpected correlation led to the discovery of the rôle played by ionisation in the maintenance of a completely gaseous state despite great density. The theoretical mass-luminosity relation, among others, was established through the intermediate consideration of mean molecular weight and of the relation between radiation pressure and total pressure. As we have already explained, radiation pressure contributes towards the total pressure, opposing and balancing the gravitational pressure exerted by the outer layers of a star; but when the extent of this contribution was calculated it was found (according to Chandrasekhar) to amount to no more than 1% for stars less than half as massive as the Sun, 3% in the case of the Sun itself, and 9% if the star's mass is twice that of the Sun. Consequently it plays only a minor rôle in the maintenance of the equilibrium of the majority of Main Sequence stars — i.e. gas pressure is practically sufficient to balance the weight of the star's material. Nonetheless, this equilibrium is described as radiative, since it is by radiation alone that a star gets rid of its surplus energy.

The foregoing account does not apply to the central core of a star, however. Investigation of the degree of stability of the radiative equilibrium has shown that this falls off progressively at increasingly deep levels; according to Cowling, the equilibrium ceases altogether to be stable at a distance from the centre equal to one-sixth of the star's radius. Within the core itself it is replaced by convective equilibrium — an equilibrium depending on the transfer of energy, not by radiation, but by convection currents in the material of the star; these may be likened to the rising currents of warmer air (thermals) in our own atmosphere. In the stellar model known as the convection nucleus, the central region (amounting to only 0.5% of the whole star by volume, or 15% by mass) generates virtually the whole of the radiant energy which ultimately finds its way to the surface and escapes into space.

Radiation pressure nevertheless plays a rôle of rapidly increasing importance as the mass of the star increases; for a mass equal to 4 times that of the Sun it accounts for 20% of the total pressure, at 6 times the Sun's mass, 30%, and at 10 times the Sun's mass, 40%. Eddington showed that this rapid increase of radiation pressure with increasing mass explained the fact that there are no stars more than about 50 times as massive as the Sun. By the time such a mass is attained, the radiation pressure has increased to a point where it overcomes the opposing force of gravity, and the star becomes unstable: blown out by the pressure of its own radiation it will automatically disintegrate into fragments whose individual masses are below the critical limit, and which will therefore be stable.

This dependence of radiation pressure upon stellar mass has a bearing on the mass-luminosity relation. The luminosity of the dwarfs, with their small masses and very weak radiation pressure, varies as the fourth power of their mass, but in the giants, with their great masses and correspondingly powerful radiation pressure, it is

* For example, carbon, with an atomic weight of 12, has 6 electrons; iron, atomic weight 56, has 28.

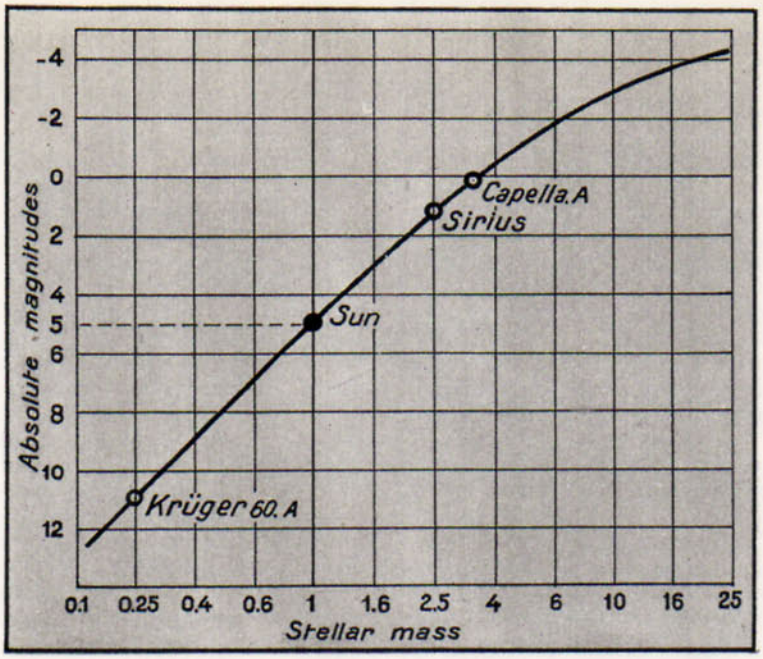


Fig. 622 — THE THEORETICAL MASS-LUMINOSITY RELATION, AS ESTABLISHED BY EDDINGTON. (Cf. Fig. 615)

directly proportional to mass. Since, moreover, there is some reason to believe that the chemical composition of the stars (in particular, their hydrogen content) also varies with their masses, it is not surprising that the empirical mass-luminosity relation (p. 348) should differ slightly from the theoretical relation, whose derivation involved certain assumptions introduced to simplify the problem. But the mere fact that the relation can be predicted theoretically is itself of the first importance.

White dwarfs and degenerate matter. — We have seen that, because of ionisation and despite its great internal density, the whole of a star behaves as a perfect gas, and also that from the supergiants to the dwarfs the stars observe Eddington's mass-luminosity relation. We also know that there is one abnormal class of stars whose members, the white dwarfs, do not abide by this relation, being relatively too massive for their luminosities; the reason for this is that their dimensions (and in particular the areas of their luminous surfaces) are abnormally small — in size, they compare with planets like the Earth rather than with the stars. We have also seen that since they compress the mass of a star into the volume of a planet their internal densities must exceed that of the densest metals by a factor of 10^4 or 10^5.

How are such densities to be accounted for? And what must the physical state of the matter comprising a white dwarf be like? The earliest answer to these questions was provided in 1926 by the English astrophysicist Fowler, with the help of the then only recently developed wave mechanics. We have seen that the matter composing the cores of normal stars continues to behave as a perfect gas, despite its density, because the atomic nuclei and free electrons occupy much less space than neutral atoms. But what happens if the whole mass of a star is compressed into a much smaller volume, as it is in the white dwarfs? At these enormous densities the heterogeneous collection of nuclei and electrons would be so tightly packed that it could no longer behave as a perfect gas. It would nevertheless constitute a gas in one sense, since consisting of independent particles endowed with random motions, but it would have properties very different from those that we commonly think of as 'gaseous'. The electrons of a neutral atom possess a minimum kinetic energy which maintains them in their orbits and prevents them from falling in to the nucleus, and when they are detached from the normal atomic structure (which is too large to exist in such a dense environment) this energy is conserved in their random motions; these motions in turn constitute the considerable internal pressure exerted by the electron gas against the downward pressure of the star's outer layers.

This electron gas — often named after Fermi, the Italian physicist who first investigated it theoretically —

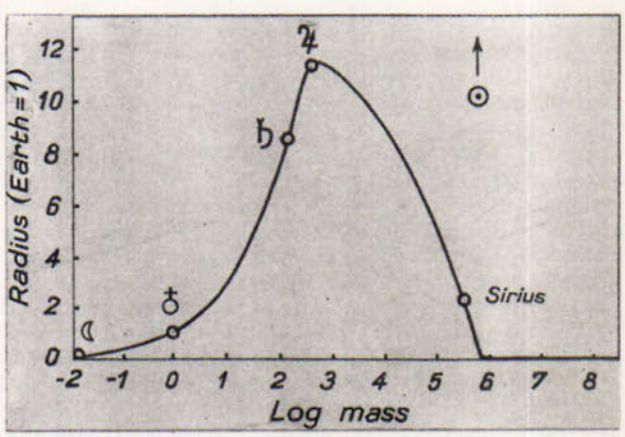


Fig. 623 — RELATION BETWEEN THE MASS (EARTH = 1) AND RADIUS OF COOL BODIES such as planets (left half of curve) and white dwarfs that have exhausted their hydrogen supply (right half) — according to the Indian physicists Chandrasekhar and Kothari. The right-hand branch of the curve corresponds to degenerate matter, where the radius diminishes while the mass increases, and in the case of very great masses finally becomes nil. The arrow indicates the present position of the Sun (R = 109).

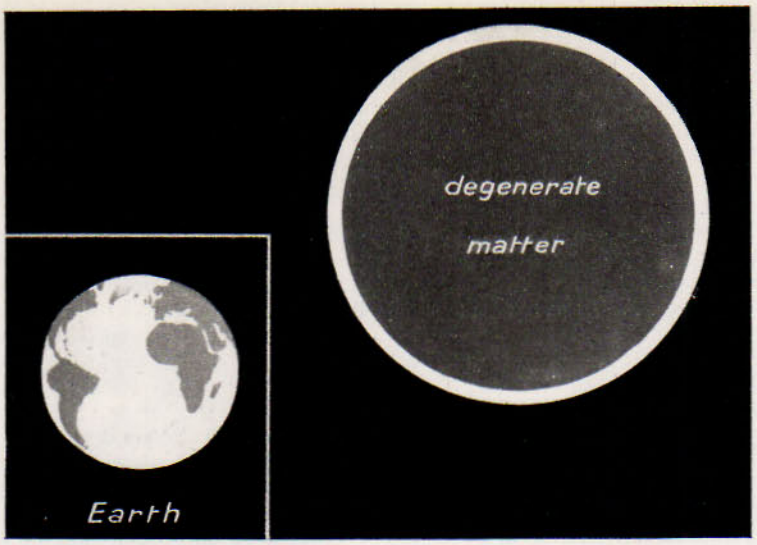


Fig. 624 — SECTION OF THE COMPANION OF SIRIUS, SHOWING ITS DARK NUCLEUS OF DEGENERATE MATTER. Only a thin superficial layer of radiating matter remains. (For comparison, the Earth is shown on the same scale.)

possesses some very odd properties, quite unlike those of a normal gas. In particular, Chandrasekhar has demonstrated that matter in this so-called degenerate state confounds common sense by occupying a *smaller* volume the *greater* its mass, since the pressure of Fermi's gas increases less rapidly than that of the star's upper layers, resulting from the overlying weight, when the total mass is increased. The pressure at which the atomic structure of normal solid matter collapses and passes into the degenerate state has been calculated by the Indian astrophysicist Kothari: it is of the order of 10^7 atmospheres. Below this pressure, matter may be very highly compressed but it retains relatively familiar properties — the larger a planet is, for instance, the more massive it is. Above the critical pressure, on the contrary, the matter collapses and becomes a degenerate gas — the smaller a white dwarf, the more massive it is.

Another difference between normal and degenerate matter is that whereas the ordinary solids of a planetary interior remain more or less homogeneous owing to their rigidity, the density of the degenerate matter increases rapidly towards the centre of a white dwarf, since it is as compressible as ordinary gas. It has been calculated that at the centre of Jupiter the pressure almost reaches the critical value at which atomic structure collapses: the giant planet thus represents the upper limit of size for a cold body composed of ordinary matter.

Another fundamental but strange property of degenerate matter is that even when its temperature is raised to the hundred million degrees mark (as it is in the interiors of white dwarfs) it can emit no radiation: to an external observer it would appear to be both dark and cold. This will apear less paradoxical if it is remembered that temperature is simply a measure of the agitation of the basic particles of matter, which in the case of electrons in a stellar interior must exceed several thousands or tens of thousands of m.p.s. if it is to provide sufficient pressure to maintain the star's equilibrium; there is therefore no doubt about the high temperature of a white dwarf core. At the same time, however, the emission of radiation by an atom occurs only when the arrangement of its orbital electrons changes from a higher to a lower energy level, the difference between the two being carried by the photon emitted during the transition (see Chapter 20); and in the case of degenerate matter such transitions are impossible because the electrons are already forced into their minimum energy states by the crushing external pressure. Furthermore, theory has shown that the number of available positions in a degenerate gas is strictly limited (Pauli's exclusion principle), and when these are all occupied — i.e. when the matter is completely degenerate — it can no longer emit radiation. A body in this limiting condition might be described as a 'black dwarf'.

The outer shell of a white dwarf, some hundreds of miles thick, does not consist of degenerate matter, since the pressures near the surface are low enough for the atoms to retain their electronic structure; it is owing to the radiation emitted by this layer that we are able to see the white dwarfs at all. It will now be understood why white dwarfs do not satisfy the mass-luminosity relation: the greater part of their mass plays no part in the production of the observed radiation, which is emitted solely by this thin surface layer.

The spectrum of this radiation is also very abnormal; often it consists of a continuous background (indicating high pressures) which is sometimes crossed by a few absorption lines, especially those of hydrogen. Despite this it is generally agreed that the degenerate interior contains little or no hydrogen; Chandrasekhar's mass-luminosity relation depends on the proportion of the atomic mass to the number of free electrons, which contributes the greater part of the internal gas pressure; this proportion is 1:1 for hydrogen, and 2:1 for helium and all the other elements. Now all the white dwarfs so far discovered lie fairly close to the theoretical curve derived from the latter value. This scarcity of hydrogen in the white dwarfs will take on an added significance when we come to examine the problem of solar and stellar evolution in Chapter 18.

But first of all we must learn something of the origin and mechanism of the energy production that goes on in the interiors of the stars.

ATOMIC ENERGY IN THE SUN AND THE STARS

Solar and stellar radiation. — The study of the Sun and the stars has provided innumerable proofs of the very high rate of energy production which is occurring all the time in their interiors. In the case of the Sun we

have already been able to make a quantitative estimation of its ceaseless output of radiant energy; some of the stars have been found to be less prodigal than the Sun, but many others are conspicuously more spendthrift. In addition, we have just seen that the source of this energy is the central regions of the stars, where the temperatures reach a level of several 10^6 or 10^7 degrees.

The existence of such temperatures, and the origin and maintenance of radiation on such a scale, for a long time posed insoluble problems, and their answer has indeed only been found during the past 25 years.

The Sun, a typical Main Sequence dwarf, is the star that we known most intimately; it is also the only one whose radiation in past epochs can be deduced from terrestrial clues. Furthermore, it was in connection with the Sun that the problem of the source of stellar radiation first arose. To start with, therefore, we shall consider the Sun rather than the other stars.

The problem is easily enough stated. Absolute measurements of the Sun's radiation have established that it transmits 2 calories to every square centimetre of the Earth's surface per minute. A simple calculation shows that the Sun's total radiation into space therefore amounts to 10^{20} (100 million billion) kilowatt-hours per second. This prodigious output of energy has been maintained for at least 4,000 million years. The so-called 'radium clock' — the relative quantities of uranium and lead in the radioactive minerals of the Earth's crust — has in fact shown that the oldest strata have been in existence for about 4×10^9 years (see Chapter 18), while the age of the earliest fossil-bearing strata, which sets a minimum time limit to the existence of life on the Earth, has been determined with sufficient accuracy as about 10^9 years. Now the phenomenon of life demands from its environment, among other things, a very restricted temperature range, so that we can say with confidence that the intensity of the Sun's radiation at that remote epoch was not materially different from what it is today.

The problem, then, is simply this: what possible mechanism can have maintained the Sun's enormous output of energy unfalteringly for at least the past thousand million years?

It is obvious, and has been realised for a long time, that the Sun cannot be 'burning', in the usual sense of the word: it is in no way analogous, on an enormously exaggerated scale, to the fire in the domestic fireplace. Even if the Sun were supposed to be composed entirely of coal, and supplied with all the oxygen necessary for its combustion (which is not the case, the Sun being isolated in empty space) it would be completely burnt out in a few thousand years.

Equally unsatisfactory was the mechanism suggested in 1848 by the physicist R. Mayer, who thought that the solar radiation might be maintained by the kinetic energy of meteorites constantly falling into the Sun. Calculation proves that the weight of meteorites that the Sun would have to receive each year would be one twenty-millionth of its present mass, which means that since the times of antiquity its mass must have increased by one ten-thousandth — a large enough increment to produce observable, but in fact non-existent, modifications of the planetary orbits and motions.

The earliest explanation that can begin to be described as scientific was that formulated by the German physicist H. von Helmholtz in 1854. He pointed out that a gaseous mass like the Sun must, in cooling, tend to contract. This contraction will convert potential into kinetic energy since it will cause the molecules to 'fall' towards the Sun's centre; this in turn will produce a rise in temperature which will counteract the radiation losses. Unfortunately, this source of energy is quite inadequate for the task it is supposed to perform: the calculations of Lord Kelvin and Henri Poincaré, about 1900, showed that if this were its sole method of energy production the Sun could not have been radiating at its present rate for more than 2.5×10^7 years. This does not even cover the Earth's geological past: the forests now buried in our coal mines were flourishing at least 10 times as long ago.

Subatomic sources of energy. — The first step towards the correct solution — though it was not recognised as such until very much later — was made in 1896 by Henri Becquerel, with the discovery of radioactivity. This presented mankind with an entirely novel phenomenon — spontaneous explosions occurring within the atom itself, with the liberation of a considerable quantity of energy. The seat of this explosive metamorphosis was soon identified as the atomic nucleus.

The researches of the New Zealand physicist Ernest Rutherford (1911) showed that atoms consist of a central, positively charged nucleus (which concentrates within a very small volume practically the whole mass of the atom), and a number of much lighter, negatively charged electrons revolving round it.

The second vital element in the solution was provided by Einstein in 1905; this, the principle of the equivalence of mass and energy, was a consequence of his special theory of relativity. It stated, simply, that energy has inertia, and conversely that any mass m is equivalent to a certain quantity of energy E; this quantity is given by the equation

$$E = mc^3$$

where c is the velocity of light (300,000 k.p.s.). It follows from this equation that 1 gm. of matter is equivalent to 2×10^{13} calories, and that in annihilating 1 gm. of matter this quantity of energy will be released.

From these data it can easily be calculated that the Sun's output of energy corresponds to the destruction of more than 4 million tons per second, or 1.3×10^{14} tons per year. Fortunately, the Sun's mass is so enormous (2×10^{27} tons) that even at this rate of loss it would only have been reduced by 0.1% during the last 1.5×10^{10} years.

This principle of the equivalence of mass and energy must not be thought of as a mathematical fiction; on the contrary, it is the expression of a fundamental physical reality, which in subsequent years was extensively confirmed both by its observed consequences and by direct experiments — notably those of the French physicists F. Joliot and J. Thibaud involving the annihilation of positive and negative electrons (1934).

If, then, this principle indicates that the Sun's release of energy represents a loss of over 4 million tons of its matter every second, then there is no doubt whatever that in its interior matter is being converted into energy at exactly this rate. Yet we still have to identify the process by which this conversion takes place.

Not long after the publication of Einstein's equivalence principle, the French physicist Paul Langevin, who was guided by it, revived a bold suggestion that had been made a century earlier (1815) by Prout, an English doctor: namely, that the various chemical elements are built up of simpler units, and that these consist of the simplest element of all — hydrogen.

In general, the atomic weights of the different elements are almost exact whole-number multiples of that of hydrogen, which is taken as unity. The few apparent exceptions to this generalisation were explained, about the same period, by Soddy's discovery of isotopes in 1910, confirmed by J. J. Thomson in 1913: an isotope of an element and the element itself have almost identical chemical and spectroscopic properties, but different atomic

Fig. 625 – NAGASAKI. 'The forces that power the Sun have been unleashed . . .' President Truman, 6 August 1945. *(Keystone photo)*

weights. The mixture of several different isotopes with the normal element can in certain cases produce atomic weights which are fractions of those expected by classical chemistry. Thus chlorine of atomic weight 35.5 is really a mixture of two main isotopes, whose atomic weights are respectively 35 and 37.

The nucleus of the hydrogen atom thus being confirmed in its status as the primary physical particle, it was christened the proton (*protos*, original) by Rutherford in 1919.

Meanwhile, F. W. Aston's accurate determinations of the atomic weights – or rather of the masses of the nuclei – of various isotopes, by means of the mass spectrograph, had revealed small residual and irreducible differences between these masses and the nearest whole multiples (atomic numbers) of the mass of the proton. Following Langevin's suggestion, he interpreted these defects of mass as a measure of the energy liberated during the synthesis of a nucleus from two or more protons, in accordance with Prout's theory and the relativistic formula for the inertia of energy (F. W. Aston's 'packing fraction', 1917).

In 1919, taking these results as his starting point, the French physicist Jean Perrin pointed out for the first time that the synthesis of helium from hydrogen could provide the Sun with an incomparably richer supply of energy than any that had previously been suggested, and would be capable of maintaining its present rate of radiation for more than 10^{11} years – a time scale capable of satisfying the demands of geologists, with plenty in hand.

Helium is the next simplest element to hydrogen, and occurs abundantly on the Sun – where, as its name suggests, it was first discovered, spectroscopically, by N. Lockyer in 1869; at that time it had not yet been isolated on Earth. Now the exact atomic weight of helium is 4.00386, while that of hydrogen is 1.00813; deducting the masses of the orbital electrons, the masses of the nuclei are respectively 4.00276 and 1.00758. If, therefore, 4 hydrogen nuclei combine to form a helium nucleus, a fraction of the unit of mass* equal to

$$(4 \times 1.00758) - 4.00276 = 0.02766$$

or 1/145 of the initial mass, is surplus to the helium nucleus, and in accordance with the equivalence principle will appear in the form of energy released by the reaction.

In the same year, 1919, Lord Rutherford showed experimentally that such atomic transmutations were not just theoretical pipe-dreams, but could be produced artificially by the bombardment of atomic nuclei with particles of various kinds – particularly protons (hydrogen nuclei) and α-particles (helium nuclei) – whose kinetic energy is sufficient to overcome the electrical repulsion of the nuclei.

The rapid progress of nuclear physics since that epoch-making experiment is a matter of common knowledge. Milestones on the route were the discovery of the neutron by J. Chadwick in 1932, of the positive electron or positron by C. Anderson in the same year, of artificially provoked radioactivity by I. and F. Joliot-Curie in 1934, and finally of the fission of uranium by O. Hahn and F. Strassman in 1939, which paved the way to the atom bomb.

Nuclear reactions in the Sun and the stars. – Thus for the past 30 years or more astronomers have been confident that the Sun and the stars derive their radiation from atomic energy liberated in their interiors. It is agreed that at the temperatures of several 10^6 or 10^7 degrees assigned to these regions by the theory of stellar radiation, the thermal agitation of the atoms** would endow them with sufficient kinetic energy to overcome their mutual repulsive forces and set off a variety of nuclear reactions.

But it was not until 1938 that atomic physics had advanced to the point when a large number of nuclear reactions could be catalogued. Among these was the chain of reactions involving the lightest and most abundant elements on the Sun, which the physicists Hans Bethe in America and C. von Weizsäcker in Germany discovered independently of one another: these explained how, in the 20 million degree furnace of the solar core, hydrogen could be converted into helium.

This chain reaction – often called Bethe's cycle, or the carbon-nitrogen cycle – can be divided into 6 stages, of which 4 (nos. 1, 3, 4, 6) are practically instantaneous transmutations, while 2 (nos. 2, 5) are radioactive disintegrations. They can be represented as follows***:

$$1.\quad C_6^{12} + H_1^1 \longrightarrow N_7^{13} + \gamma$$

$$2.\quad N_7^{13} \longrightarrow C_6^{13} + e^+$$

* The physicist's unit of atomic mass is defined by giving the commonest isotope of oxygen an atomic weight of 16.0 precisely. Its value is 1.66×10^{-24} grams.

** More correctly, of their nuclei, since the atoms will at these temperatures be stripped of practically all their electrons, as explained on p. 352.

*** In these formulae, the symbols H, He, C, N and O represent respectively the nuclei of hydrogen, helium, carbon, nitrogen and oxygen, $e+$ represents a positive electron, and γ represents photons of radiation. The symbols on the left-hand side are the reacting particles, and those on the right the particles resulting from the reaction. The inferior indices are the atomic number (nuclear charge) and the superior indices the atomic weight (mass of the nucleus).

$$3.\quad C_6^{13} + H_1^1 \longrightarrow N_7^{14} + \gamma$$

$$4.\quad N_7^{14} + H_1^1 \longrightarrow O_8^{15} + \gamma$$

$$5.\quad O_8^{15} \longrightarrow N_7^{15} + e^+$$

$$6.\quad N_7^{15} + H_1^1 \longrightarrow C_6^{12} + He_2^4$$

$$\text{Net result: } 4\,H_1^1 \longrightarrow He_2^4 + 2e^+ + \gamma$$

Throughout this cycle carbon plays a rôle analogous with that of a catalyst in chemical reactions: at the end it is regenerated, and ready to perform the same rôle again. During the course of the cycle, also, 4 hydrogen nuclei have disappeared and 1 helium nucleus has been formed, with the liberation of energy equivalent to the difference in their masses, as calculated above.

The energy produced by the cycle — which on the average lasts from 2 to 3 million years — amounts to

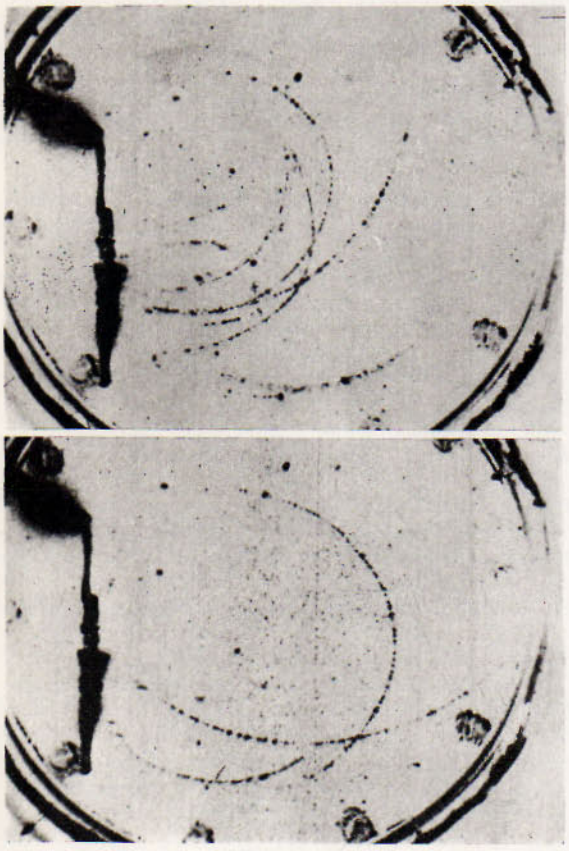


Fig. 626 — ONE OF THE NUCLEAR REACTIONS OF THE CARBON CYCLE PHOTOGRAPHED IN THE LABORATORY.
These photographs, taken with a Wilson cloud chamber by the American physicists Anderson and Neddermeyer in 1934, show a carbon plate (left) which has been bombarded with protons. The radioactive atoms of nitrogen 13 (reaction No. 1 of the cycle) break down spontaneously into carbon 13, with the liberation of positrons (reaction No. 2). The curved lines in the photographs are the paths of these positrons, their curvature being caused by a magnetic field. The lower photograph, taken 20 minutes later than the upper, shows a marked decrease of activity, indicating that fewer atoms of radioactive nitrogen remain in the carbon plate.

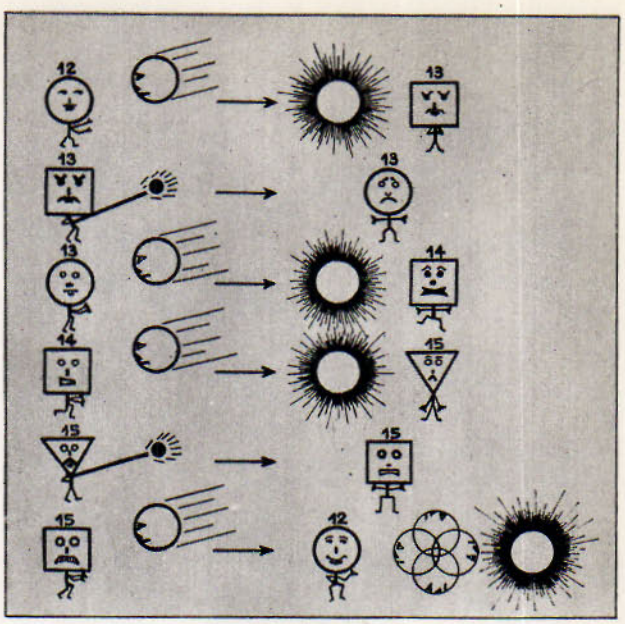


Fig. 627 — THE CARBON-NITROGEN CYCLE AS VISUALISED BY THE AMERICAN ASTRONOMER CECILIA PAYNE-GAPOSCHKIN, OF HARVARD OBSERVATORY. (Compare with the nuclear reactions in the text.)

4×10^{-4} ergs per helium nucleus produced. This does not seem very much, but the quantities of solar material involved are so immense that the cycle* produces energy at a mean rate of about 2 ergs per gram per second (25 to 50 times as fast at the Sun's centre); this computed rate agrees closely with the figure derived from observation. Other possible reactions result in energy outputs a million or more times too great or too small, and can therefore be eliminated from consideration.

There is now no doubt that it is in this way — by transmuting 564 million tons of hydrogen into 560 million tons of helium every second — that the Sun produces in its central regions the 4 million tons of light that its surface radiates per second.

What becomes of the helium that is formed? Cannot it also be transmuted, with the release of more energy? The latter question cannot be answered with a confident negative, but all experimental work in nuclear physics has gone to show that the helium nucleus, or α-particle, is one of the most stable particles in nature, and it can only enter into reactions if the temperature is at least several 10^8 degrees C, which is far higher than that of the Sun's core. Helium may therefore be regarded as the dead 'ash' formed by the nuclear combustion that provides the Sun with its radiation.

At this point we should, perhaps, discuss the future history of the Sun, but the treatment of this question must wait till Chapter 18, since its significance can only be fully appreciated against a background of the modern theories of stellar evolution.

The carbon-nitrogen cycle accounts satisfactorily for the radiation output of all the stars of the Main Sequence from type M to type B or O, and the general validity of the mass-luminosity relation shows that the relative abundance of carbon and nitrogen remains essentially constant throughout this series; in the Sun, according to Russell, it is 1% (c.f. hydrogen 51%, helium 42%, and other

* The total energy produced in a given time is limited by the slowest link in the chain, i. e. No. 4, the capture of protons by the nuclei of ordinary nitrogen.

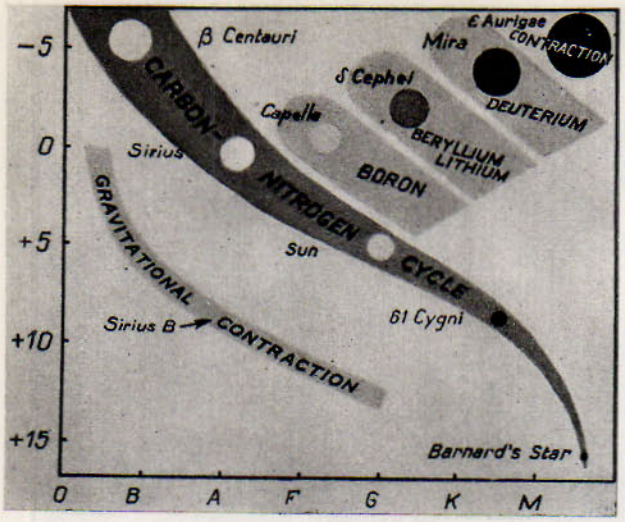


Fig. 628 — APPROXIMATE AREAS OF THE HERTZSPRUNG-RUSSELL DIAGRAM WHERE THE PRINCIPAL NUCLEAR REACTIONS — THE SOURCE OF STELLAR ENERGY — OCCUR. *(After G. Gamow)*

elements 6%). This would seem to be reasonable enough, since carbon and nitrogen emerge from the chain reaction unchanged, and would therefore be conserved indefinitely.

However, the central temperatures of the coolest stars of the Main Sequence (type M red dwarfs), as derived from the theory of radiative equilibrium, would appear to be too low for the operation of the carbon-nitrogen cycle: at a temperature of 8 to 10 million degrees C the kinetic energy of the majority of the nuclei would be incapable of overcoming the nuclear repulsion, and the activity of the cycle would be very much reduced. Bethe has suggested a different chain reaction, which can operate effectively at the lower temperatures of these stars. During the course of this cycle two protons combine to form a deuterium (heavy hydrogen) nucleus; in collision with another proton this becomes an isotope of helium; this in turn combines with an α-particle (normal helium nucleus) to form first beryllium and then lithium; finally the latter combines with another proton, and splits into two equal fragments, which are helium nuclei:

$$1.\quad H_1^I + H_1^I \longrightarrow H_1^2 + e+$$

$$2.\quad H_1^2 + H_1^I \longrightarrow He_2^3$$

$$3.\quad He_2^3 + He_2^4 \longrightarrow Be_4^7$$

$$4.\quad Be_4^7 + e- \longrightarrow Li_3^7$$

$$5.\quad Li_3^7 + H_1^I \longrightarrow 2\ He_2^4$$

Although this chain of reactions can occur at a lower temperature than the carbon-nitrogen cycle, its beginning and end points are the same — hydrogen is converted into helium.

In the same way other reactions have had to be found, which are capable of operating at lower and lower temperatures, in order to account for energy production along the giant branch, where the central temperatures fall ultimately to a few hundred thousand degrees C. Of these reactions, the easiest to start is that producing helium by the collision of a proton and a deuterium nucleus:

$$H_1^2 + H_1^I \longrightarrow He_2^3$$

This can occur at a temperature of only 500,000°, and is therefore possible in the interiors of red giants and type M long-period variables. At a slightly higher temperature, of the order of 2×10^6 degrees (e.g. type K giants) lithium becomes reactive, and under bombardment by protons again produces helium nuclei:

$$Li_3^7 + H_1^I \longrightarrow 2\ He_2^4$$

$$\text{and } Li_3^6 + H_1^I \longrightarrow He_2^4 + He_2^3.$$

In the G and F type giants — and, particularly, in the classical Cepheids, whose central temperatures range from 3 to 9 million degrees — the following reactions involving beryllium and boron can take place, again with the production of helium:

$$Be_4^9 + H_1^I \longrightarrow Li_3^6 + He_2^4$$

$$B_5^{11} + H_1^I \longrightarrow 3\ He_2^4$$

The lithium formed in the first of these reactions is, as before, rapidly transmuted into helium by collision with a proton:

$$Li_3^6 + H_1^I \longrightarrow He_2^4 + He_2^3$$

Finally, in the type A giants and cluster variables, whose central temperatures are still higher, the reaction

$$B_5^{10} + H_1^I \longrightarrow C_6^{11}$$

(in which a rare isotope of carbon is formed from the normal isotope of boron) may also contribute to the stars' energy production. But at this point we rejoin the Main Sequence, where the carbon-nitrogen cycle is capable of accounting for the greater part of the radiation.

A word must still be said about the white dwarfs. We know that these stars contain practically no hydrogen, and that what little there is is confined to their surface layers. Indeed, as Gamow has pointed out, if they contained considerable quantities of hydrogen it would react violently at the high temperatures of these stars (several million degrees even in the non-degenerate surface shell) and they would accordingly be very much brighter than they are in fact observed to be. It seems, therefore, that in this case Helmholtz's theory has been vindicated, and the white dwarfs shine at the expense of their gravitational energy, liberated by continuous contraction. However, it has to be admitted that their nature is still too imperfectly understood for any final conclusions to be drawn.

Despite this, the data already acquired on the sources of stellar energy enable us, as we shall see in Chapter 18, to build up a fairly coherent picture of the course of stellar evolution, which unites (at any rate provisionally) a wide range of miscellaneous facts.

Before embarking on this intriguing (if still, very often, hypothetical) subject of cosmogony, we must continue our survey of the known facts concerning the Universe as it is at present.

Fig. 629 — NGC 6960, THE FILAMENTOUS NEBULA IN THE CYGNUS REGION OF THE MILKY WAY, WHICH SEPARATES TWO STAR FIELDS OF UNEQUAL DENSITY. *(Photo, Mount Wilson Observatory, 100-inch telescope, exposure 7 hours; 1921)*

CHAPTER 16

THE GALAXY

The preceding chapters have introduced us to the teeming multitude of the stars, to their many and varied individual characteristics and also to their underlying similarity. Among these stellar hosts our Sun cuts a rather modest figure. We have also learnt that the stars are separated from one another by distances that are to be numbered in tens, hundreds or thousands of lightyears, and that their numbers increase steadily from magnitude to magnitude; the largest telescope at present in existence shows something like a thousand million stars. The question immediately suggests itself: will every future improvement in our means of observation reveal more and more stars, as hitherto, or will an end eventually be reached? In other words, does the universe of the stars form a finite system whose frontiers we might one day reach? We now know that the system is finite, and we shall see how the extent, frontiers and structure of the gigantic stellar system to which our Sun belongs — the Galaxy — have been gradually revealed.

The Milky Way. — On a summer evening, the reader has no doubt often admired the magnificent luminous band — ill-defined, mysterious, infinitely complex in its structure — which stretches above his head from horizon to horizon, and which we call the Milky Way. Rising from the north-east horizon, it passes through the constellations Perseus, Cassiopeia and Cepheus, becoming noticeably brighter in Cygnus where it splits into two unequal branches, and thence crossing Scutum and Ophiuchus, to disappear below the south-west horizon by way of the constellations Scorpio and Sagittarius. Beyond the horizon it continues across the southern skies, where, indeed — in the constellations Centaurus, Crux and Carina, Vela and Puppis* — it is to be seen in its full magnificence; further on still, and now much less brilliant and more regular, it passes through our northern winter constellations of Canis Major, Orion, Taurus and Auriga, to come back once more to Perseus. Thus the Milky Way spans the whole heavens, varying in brightness, regularity and width, but following roughly the course of a great circle around the star sphere.

Owing to the fundamental rôle that this great circle plays in all investigations of the structure of the stellar universe, it has been made the basis of a system of co-ordinates — galactic longitude and latitude — of which it is the equator. Galactic latitudes run from 0° (the galactic equator, or median line of the Milky Way) to +90° in the northern galactic hemisphere, and from 0° to −90° in the southern; the north galactic pole is situated at R.A. 12h 40m, Dec. +28°. Galactic longitudes are reckoned from 0° to 360°, starting from the point in Aquila where the celestial and galactic equators intersect. The following are the approximate galactic longitudes of some representative stars: α Aquilae (Altair) 15°, α Cygni (Deneb) 50°, α Aurigae (Capella) 130°, α Canis Majoris (Sirius) 195°, the Southern Cross (Crux) 270½, α Scorpii (Antares) 320°.

The brightest section of the Milky Way is the southern half lying between Argo and Centaurus (galactic longitudes 240° to 60°, through 360°), where it is considerably

* See constellation maps, pp. 310—311.

more spectacular than in the hemisphere from Cygnus to Argo (galactic longitudes 60° to 240°, through 180°).

In the same way, the width of the Milky Way — established by photometric determinations of its apparent brightness — varies considerably: it is greatest (about 40°) in Scorpio and Sagittarius between longitudes 300° and 360° approximately; in the opposite region of the star sphere, Auriga and Taurus, both its width and its brightness are much reduced, and it is therefore very much less striking.

Such are the main characteristics of the Milky Way as seen with the naked eye. More detailed studies have not materially altered these general outlines, though they have made them more precise, and in so doing have provided the key to the structure of our stellar system.

The distribution of the stars. — The smallest telescope, or even a good pair of binoculars, will reveal the true nature of the Milky Way, resolving its pale pearly light into a multitude of faint stellar points; some idea of their incredible numbers is given by the plates on this and the opposite page. The luminous band that we call the Milky Way is thus built up by the accumulation of myriads of stars, too faint to be seen individually with the naked eye.

When once this had been established, it was natural that studies of the distribution of the stars on the star sphere should be initiated, and also of the manner in which they are concentrated towards the axis of the Milky Way. The earliest investigations of stellar distribution with relation to the Milky Way were made towards the end of the 18th century by William Herschel, the famous Hanoverian astronomer who made his home in England. By counting the number of stars that were visible in the field of his telescope when it was directed towards different regions of the heavens he showed that the apparent stellar density increases enormously and progressively from the galactic poles to the Milky Way, and also that this increase is much more marked in the case of faint stars than of bright. Thus although the naked-eye stars appear to be distributed more or less at random over the whole star sphere, telescopic stars are strongly concentrated towards the galactic equator, and the degree of this concentration is inversely proportional to their brightness.

The galactic concentration of the stars is strikingly illustrated in the plate on the preceding page, which shows the Milky Way and the regions bordering it between Cygnus and Cassiopeia, the field measuring 60° × 30°. Disregarding local irregularities, this concentration can be defined by means of the ratio c between the mean number of stars per unit area (say, 1 square degree) on the galactic equator and the number in the same area in the neighbourhood of the galactic poles. The table given below, due to the American astronomer Seares, quotes the mean number of stars per square degree down to various limiting apparent photographic magnitudes (m) at different galactic latitudes (b), and the corresponding value of the ratio c, the galactic concentration.

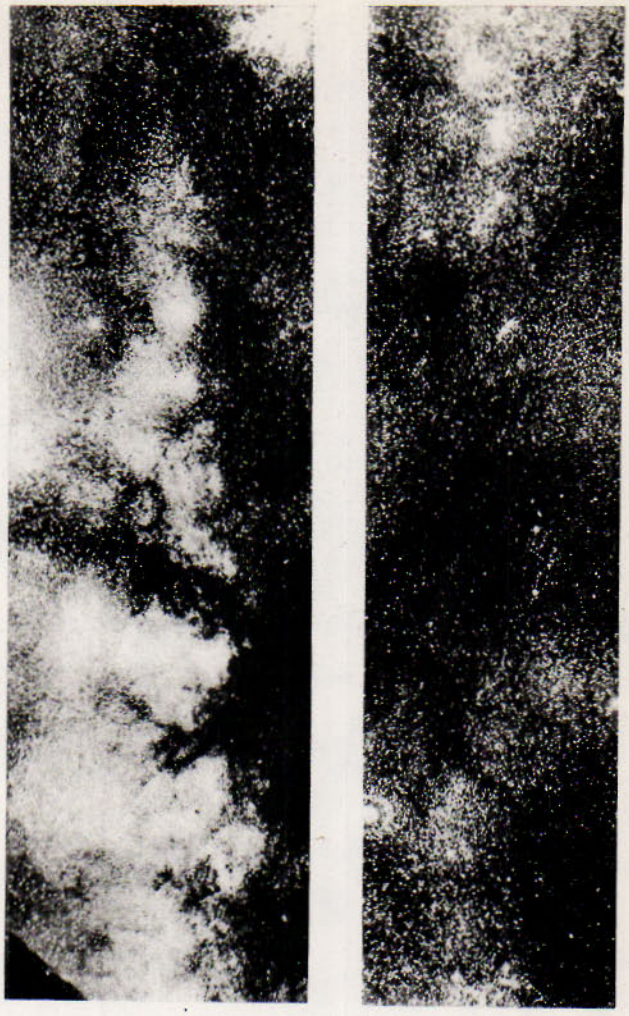


Fig. 630 — PHOTOGRAPHS PAIRED TO SHOW THE UNEQUAL BRIGHTNESS OF THE MILKY WAY TOWARDS THE CENTRE (SAGITTARIUS TO CYGNUS), LEFT, AND TOWARDS THE ANTICENTRE (CASSIOPEIA TO AURIGA), RIGHT. *(Photos, L. Rudaux)*

It will be seen that c, which is small and more or less constant for stars bright enough to be seen with the naked eye, increases rapidly with increasing m among the fainter stars, finally becoming very large indeed: there are nearly 50 times as many magnitude 21 stars per unit area in the vicinity of the galactic equator as there are at the poles.

These results can be simply explained on the assumption — made by Herschel (and suggested earlier, in the mid-18th century, by Immanuel Kant, the German philosopher, and by the Englishman, Thomas Wright) — that the whole congeries of stars forms a vast, flattened system, inside which is situated the Sun. It can easily be seen from Fig. 633 that under these conditions many more

m \ b	0°	10°	20°	30°	40°	60°	90°	c
4	0.015	0.013	0.010	0.007	0.006	0.005	0.0045	3.4
6	0.128	0.107	0.080	0.061	0.050	0.042	0.037	3.4
8	1.012	0.834	0.617	0.482	0.392	0.325	0.278	3.6
10	7.71	6.14	4.43	3.49	2.84	2.23	1.81	4.3
12	55.6	42.53	29.7	22.6	17.9	12.8	9.9	5.6
14	370.7	271.0	175.8	125.0	94.0	61.0	44.3	8.4
16	2138	1552	873	561	396	236	163	13.2
18	10187	7481	3622	1986	1313	733	482	21.1
20	40087	28973	12415	5623	3404	1824	1164	34.4
21	73620	52120	21183	8690	5011	2648	1667	44.2

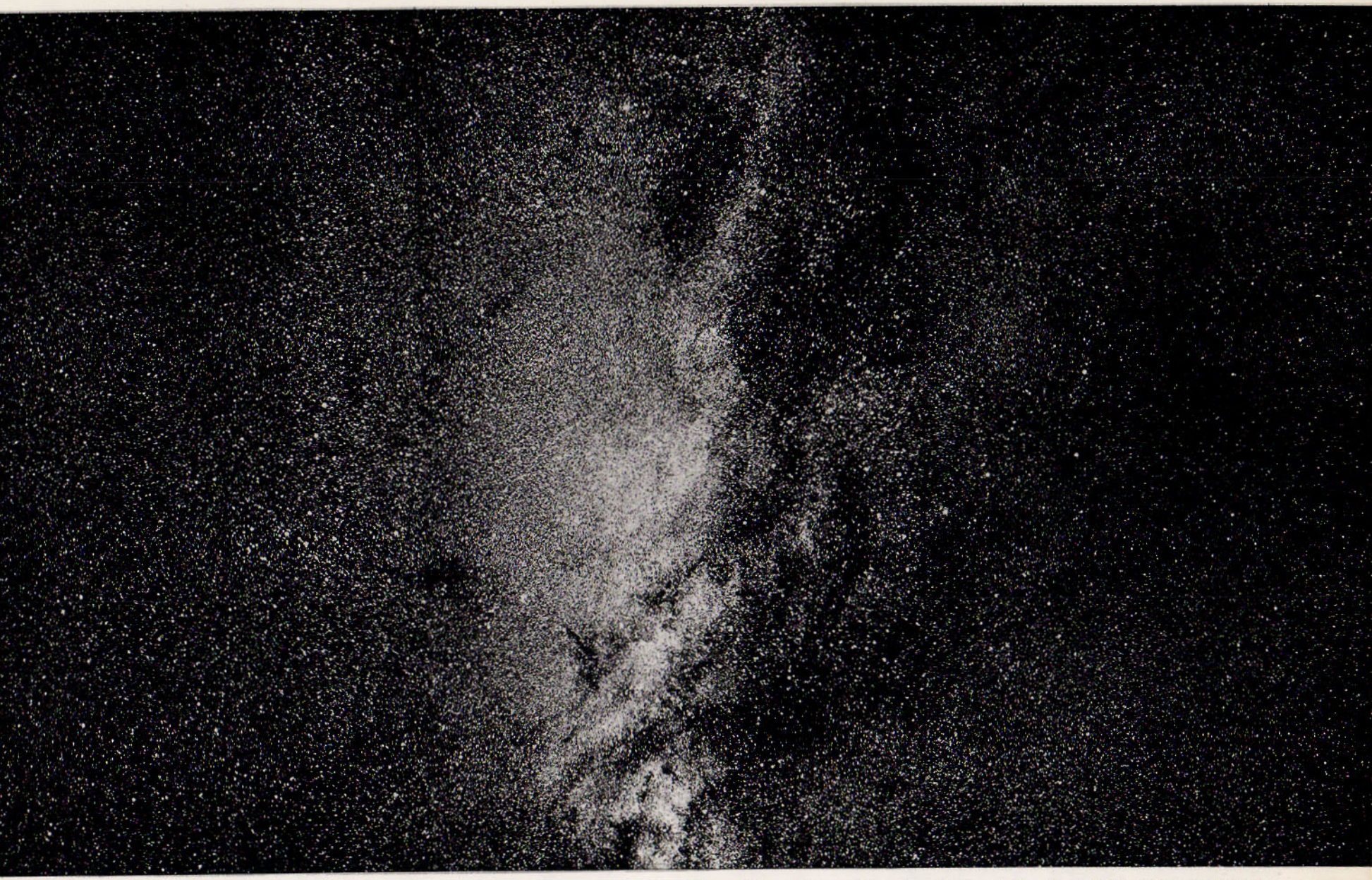


Fig. 631 — GALACTIC CONCENTRATION. This montage of three photographs covers the Milky Way and its outlying regions from Cygnus to Cassiopeia (60° × 36°), and shows the steep increase of stellar density in the vicinity of the galactic equator. *(Photo, de Kérolyr, Forcalquier Observatory)*

stars will be visible in every direction within the central plane of the system than in the two directions at right angles to it. It will also be clear why galactic concentration is increasingly strong among fainter and fainter stars: for when very large numbers of stars are considered, apparent brightness will in the mean be inversely proportional to distance; hence stars brighter than given limiting magnitudes will be contained within concentric spheres, centred on the Sun, whose diameters are proportional to these magnitudes (therefore inversely proportional to the brightness of the stars). It follows that the stars which are bright enough (therefore near enough) to be contained within a sphere whose diameter is smaller than the thickness of the Galaxy in the Sun's vicinity will be more or less equally numerous in all directions; on the other hand the fainter ones, the diameters of whose containing spheres are larger than this thickness, will be more numerous within the central plane than at right angles to it. The greater the difference between the diameter of a given sphere and the thickness of the Galaxy, the greater will be the discrepancy between the numbers of the stars brighter than the corresponding limiting magnitude in these two directions. (This assumes, of course, that the diameter of the sphere containing stars of the 21st magnitude – the faintest that can at present be reached – is smaller than the extension of the Galaxy in its equatorial plane.)

Herschel at first assumed that the Sun lies near the centre of the stellar system, but later investigation showed that this attribution of a privileged position was without justification. The position of the Sun within the Galaxy is in fact quite far from the centre; even a naked-eye view of the Milky Way is enough to show that the stars are much more numerous in the direction of Sagittarius than in the opposite direction, towards Auriga. Furthermore, the laws of mechanics, as well as the example of almost all the other stellar systems that we are able to observe, suggest that the Galaxy is strongly concentrated towards its centre, with a comparatively low star density in its outlying regions. This spatial concentration towards the centre will tend to accentuate the difference between the Milky Way's brightness in the direction of the galactic centre and that of the poles.

It can also be proved quite easily that the stars are not only concentrated towards the centre of the Galaxy, but also towards its central plane. The table on p. 367 shows that even the brighter stars are markedly concentrated towards the Milky Way: does this mean, then, that the sphere containing them already stretches beyond the edges of the Galaxy on either side of its central plane?

Fig. 632 – THE MILKY WAY FROM AQUILA TO SAGITTARIUS; CENTRE, THE GREAT STAR CLOUD IN SCUTUM. *(Photo, de Kérolyr, Forcalquier Observatory)*

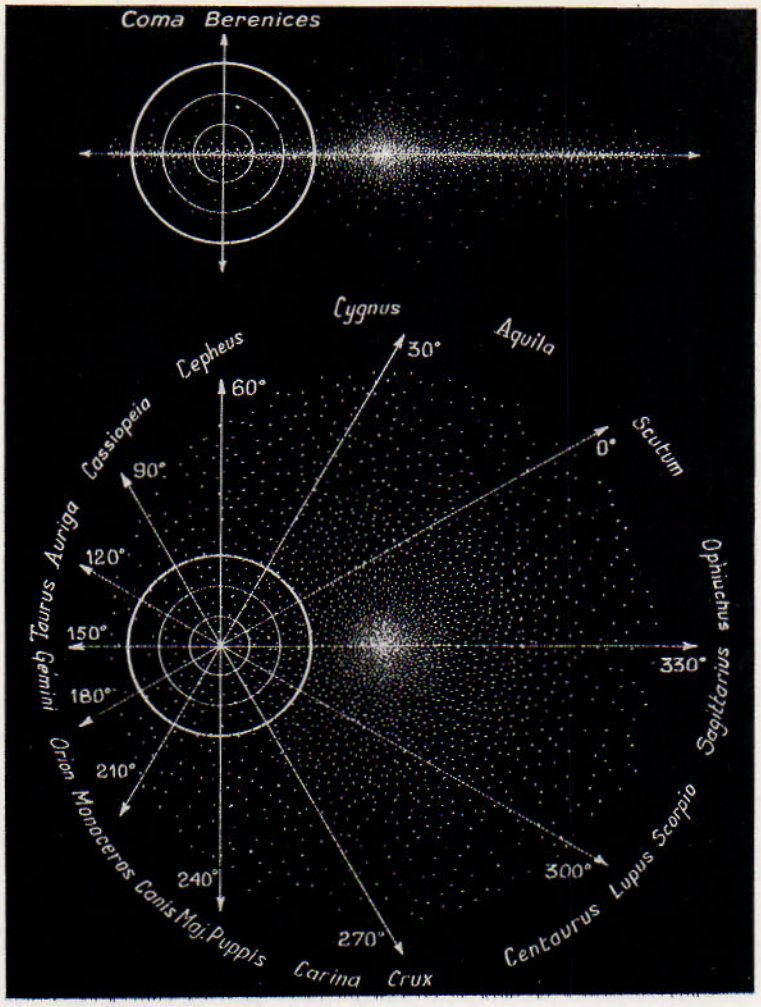


Fig. 633 — DIAGRAM OF THE GALAXY (above, in section; below, in plan), explaining the galactic concentration of the stars and its variation as a function of galactic longitude.

Obviously not, because we can also see that the value of c hardly changes between $m = 4$ and $m = 10$ — yet stars of the tenth magnitude are, on the average, 10 times as distant as those of the fourth. In the case of the brighter stars, therefore, the observed galactic concentration is not merely an effect of perspective, as in the case (largely) of the fainter ones, but must result from real spatial crowding in the neighbourhood of the galactic plane, with progressive thinning-out on either side of it.

The number of stars in the Galaxy. — Once the general plan of the Galaxy is established, the next question is the total number of stars that comprise it, and the manner in which they are distributed. A first approximation to the complete answer is provided by star counts. We have seen that the total number of stars increases regularly in proportion to their increasing faintness; but although this increase is continuous it appears to fall off progressively as fainter and fainter stars are taken into consideration. This can be accounted for by considering the way in which the relation between the total number of stars down to successive magnitude limits varies as a function of magnitude. The importance of this relation, known as the magnitude function, lies in the fact that if the stars were all similar to one another and distributed uniformly in space its value would be 3.98, whereas in fact it does not exceed 3.0 for the brightest naked-eye stars, and decreases steadily from magnitude to magnitude till at the 21st its value is only 1.5. If this trend is extrapolated, it can be established at what magnitude its value would become zero, i.e. the magnitude of the faintest star if we could see all the stars in the Galaxy. And from this, the total number of these stars can be derived.

Seares has calculated from his star counts that the total number of stars in the Galaxy is about thirty thousand million, and that the apparent magnitude of the faintest of these is 60. But it must be appreciated that such an extrapolation as this is a risky business, for it is based on statistics of stars to magnitude 20 only, and these number less than 1,000 million or hardly 3% of the derived total. Seares's figure must therefore be taken for what it is worth — a rough indication of the order of size — and we must not be surprised if other methods of attacking the problem are found to produce widely different results.

In view of what we have already learnt about the enormous range of stellar luminosity and also about the galactic concentration of the stars, it is not surprising that the empirical magnitude function should be considerably smaller than the theoretical. However, even when — instead of lumping together stars of every kind over the whole sky — the star counts are restricted to stars of one spectroscopic type (therefore of more or less the same absolute magnitude), it is still found that in the galactic

Fig. 634 — THE MILKY WAY IN OPHIUCHUS. Near the centre can be seen the S-shaped dark nebula shown on a larger scale in Fig. 643. *(Photo, Barnard, Yerkes Observatory, 10-inch Bruce telescope)*

Fig. 635 — THE GREAT NEBULA IN ORION IS TYPICAL OF THE GASEOUS, FLUORESCENT GALACTIC NEBULAE. *(Photo, K. U. Mayall and H. W. Babcock, Lick Observatory, 36-inch telescope; 1939)*

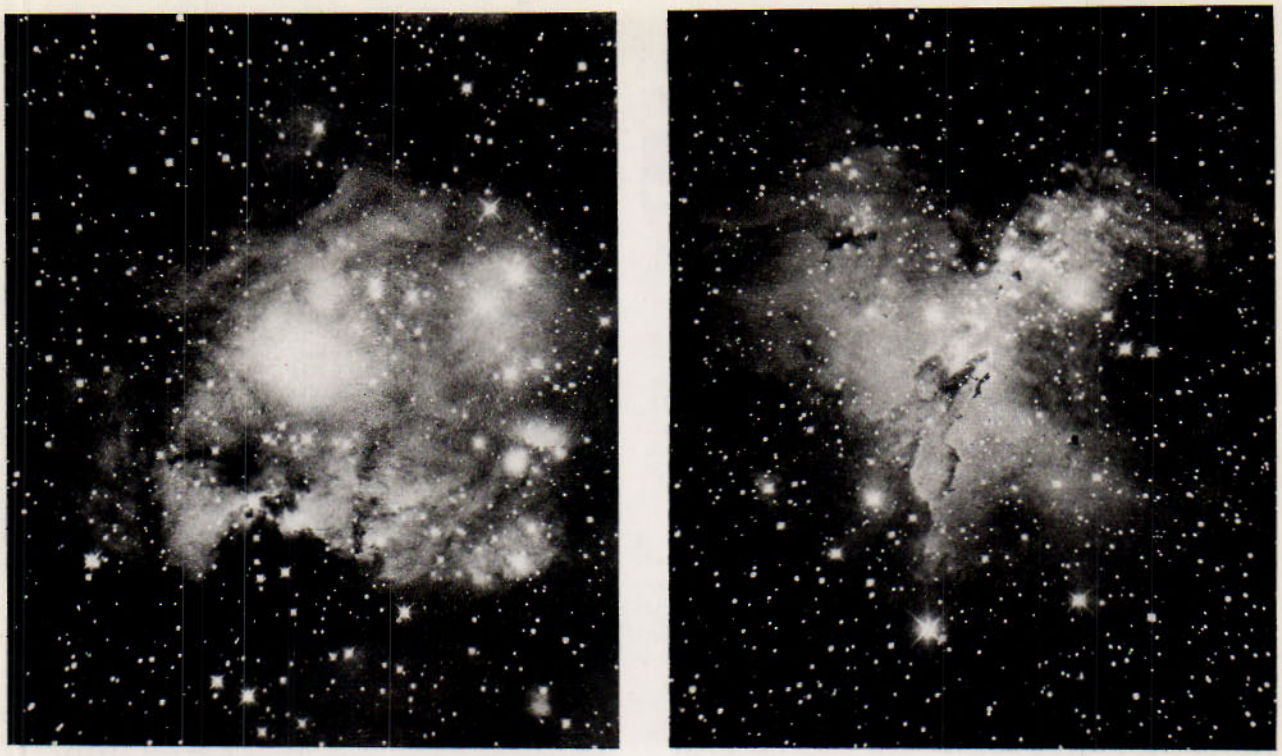


Fig. 636 — THESE PHOTOGRAPHS, TAKEN, WITH THE 60-INCH MOUNT WILSON REFLECTOR, ILLUSTRATE THE ASSOCIATION OF BRIGHT DIFFUSE NEBULAE WITH GROUPS OF BRIGHT STARS. Left: NGC 2175 in Orion, exposure 5h 10m (1921); right: NGC 6611 in Scutum, exposure 3h 25m (1919).

plane the value of the function is less than 3.98. For example, here are the results of a census of all B8 — Ao stars (absolute magnitude, +0.5) taken at Harvard College Observatory in a homogeneous region of the Milky Way measuring 100 square degrees, slightly to the north of Sirius.

APPARENT MAGNITUDE LIMIT	NUMBER OF STARS BRIGHTER THAN THIS LIMIT	MAGNITUDE FUNCTION (THEORETICAL = 3.98)
8	45	
9	141	3.14
10	339	2.40
11	1070	3.16

Thus the number of stars is, on the average, barely tripled from one magnitude to the next, whereas theoretically it ought to be nearly quadrupled.

This deficiency is encountered in all similar star counts in the Milky Way region, and can be explained in two ways: either the stars really do thin out in every direction from the Sun, in which case the latter must be at the centre, not of the Galaxy (for we shall see that there are convincing reasons for believing that this is not so) but of a local stellar condensation — the so-called Local Cluster; or else interstellar space is not perfectly transparent, so that light reaching us from distant stars is weakened, making them appear more distant still, and therefore distributed throughout a larger volume of space than they really are. We shall see later that although the existence

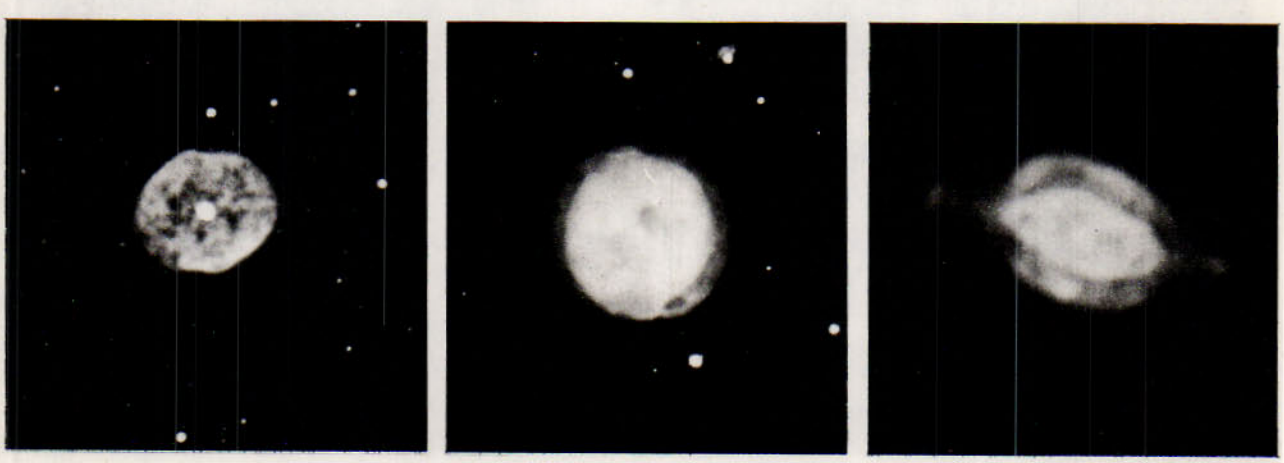


Fig. 637 — THREE INTERESTING PLANETARY NEBULAE. Left: NGC 1501 in Camelopardus, exposure 2h (1913); centre: the Owl Nebula, M 97, in Ursa Major, exposure 4h (1910); right: the Saturn Nebula, NGC 7009, in Aquarius, exposure 3h 30m (1912). *(Photos, G. W. Ritchey, Mount Wilson Observatory, 60-inch reflector)*

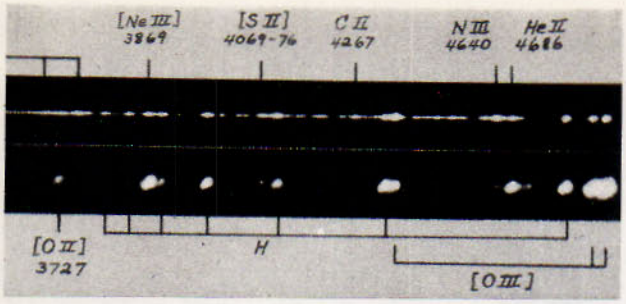


Fig. 638 — SIMILARITY BETWEEN THE SPECTRUM OF NOVA HERCULIS ON 4 JUNE 1935 (SIX MONTHS AFTER ITS OUTBURST), ABOVE, AND THAT OF A PLANETARY NEBULA, NGC 7027, BELOW. Note the 'forbidden' lines of ionised oxygen, O II, O III, of ionised nitrogen, N III, and of ionised helium, He II, which constitute the once mysterious 'nebulium' spectrum. The radiations marked G are those of the Balmer series of hydrogen. *(Photo, Lick Observatory)*

of a Local Cluster can by no means be ruled out, it has been definitely established that absorbing matter is widespread in interstellar space. The far from negligible effects of this interstellar material constitute one of the main difficulties that beset astronomers in their efforts to explore and chart the Galaxy.

Bright galactic nebulae. — The existence of bright nebulosity, extending through vast volumes of space, has been recognised for a long time. The nebulae are the first witness to the fact that the interstellar spaces are not void.

Some of the brighter galactic nebulae are spectacular objects when they swim into the field of a large telescope. One of the finest, and certainly the best known, is the great Orion nebula, which is easily seen with even the smallest telescope (Fig. 635). Other remarkable examples occur in the Pleiades, Sagittarius, and elsewhere. The photography reveals large areas which, because of contorted and convoluted appearance. In many cases photography reveals large areas which, on account of their faintness, escape visual observation altogether.

Fig. 639 — THE RING NEBULA, M 57, IN LYRA, a typical planetary nebula. *(Photo, Mount Wilson Observatory, 100-inch reflector, exposure 1h; 1921)*

The spectroscope has shown that in every case these nebulae are dependent for their luminescence on the radiation of nearby stars. The late Edwin Hubble, 35 years ago, discovered that their radiation is of two different sorts, depending on the temperature of the star, or stars, involved. If the star is very hot (type O, Bo or B1, exceeding 18,000° to 20,000° C) the nebular spectrum consists of a small number of bright lines superimposed on a fainter continuous background; these lines, characteristic of very rarefied radiating gas, are produced mainly by the neutral or ionised atoms of hydrogen, oxygen, nitrogen and helium, and a fair proportion of them are so-called forbidden lines (already referred to in connection with the solar corona); their intensity is an indication of the extremely low density of the source. Two of these

Fig. 640 — THE PLANETARY NEBULA M 27 IN VULPECULA, ONE OF THE LARGEST (IN ANGULAR MEASUREMENT) IN THE SKY. Note its almost perfectly symmetrical structure and its central star. *(Photo, F. E. Ross, Mount Wilson Observatory, 60-inch reflector, exposure 1h; 11 July 1931)*

Fig. 641 — DUST CLOUDS RENDERED VISIBLE BY THE LIGHT OF STARS IN THE PLEIADES, giving the impression of being swept by violent winds. *(Photo, de Kérolyr, Forcalquier Observatory, 30-inch telescope)*

lines in particular — situated in the green region of the spectrum, and responsible for the characteristic tint of the gaseous nebulae — for a long time baffled the astrophysicists, and it was supposed that these nebulae contained an unknown gas to which the name 'nebulium' was given. It was only in 1927 that the American astrophysicist, I. S. Bowen, succeeded in identifying these two mysterious radiations as emanating from doubly ionised oxygen. This degree of ionisation is produced by the very intense ultra-violet radiation emitted by the high-temperature star involved; the recombination of the ions and free electrons produces the radiation which we observe. The bright nebulae in fact shine by a type of fluorescence: the invisible ultra-violet radiation of the star is converted, via the excitation of the nebular atoms, into visible radiation. The great Orion nebula, situated some 1,000 light-years from the Sun, is one of the finest examples of this type of fluorescent nebula.

The spectra of the planetary nebulae — about which we have already said something, in connection with novae — place them in the same category. The name 'planetary' nebula was given to them in the early days of telescopic observation, because they appear as small, round or oval discs of bluish-green light, which resemble the distant planets Uranus and Neptune when viewed in small telescopes. But apart from this vague and superficial resemblance they have nothing whatever in common with planets.

When examined with large telescopes — or, better still, photographically — they assume the appearance of a fairly regular, round or oval envelope, or sometimes of a ring, at whose centre a faint blue star is invariably to be found; this star is responsible for the visibility of the nebula. The mechanism of light production in planetary nebulae is thus analogous to that already described in connection with the diffuse gaseous nebulae; they differ from these in the fact that the central star is generally even hotter than those involved in the diffuse nebulae — probably attaining or even exceeding 100,000°C.

Measurements of the diameters of planetary nebulae have shown that they are slowly expanding, which suggests very strongly that the central star is responsible not only for their illumination but also for their formation — that the star was once a nova, and that the gases

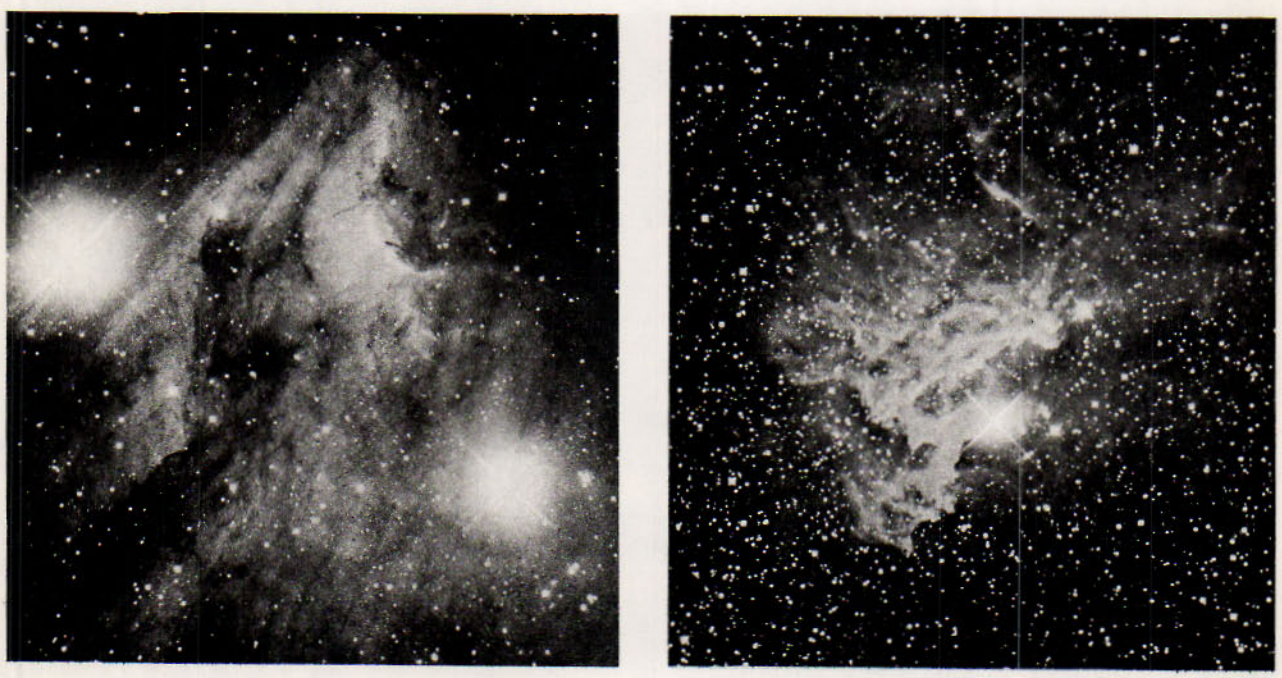


Fig. 642 — TWO BEAUTIFUL NEBULAE SHINING BY REFLECTED LIGHT. Left: the Pelican Nebula in Cygnus, exposure 8h (1933); right: IC 405 in Auriga, exposure 13h (1933). *(Photos, de Kérolyr, Forcalquier Observatory, 30-inch telescope)*

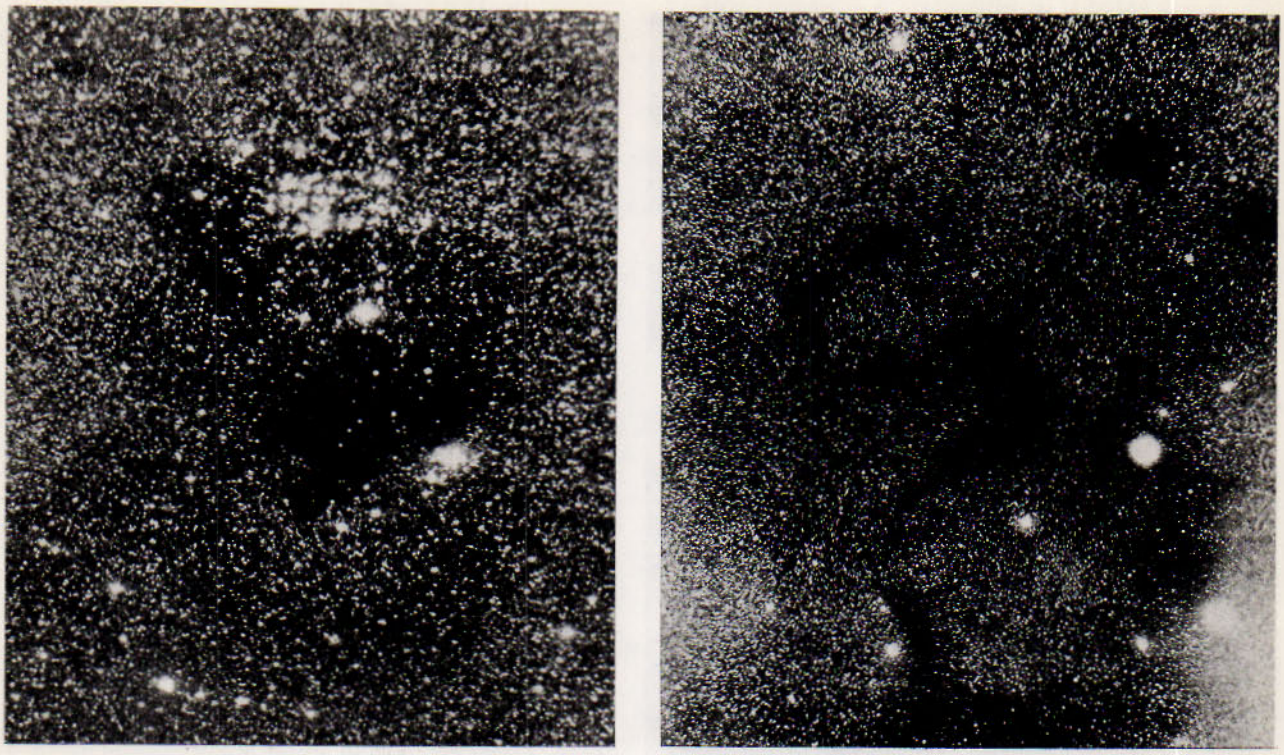


Fig. 643 — TWO FINE DARK NEBULAE. Left: Barnard 86; right: the S-shaped nebula, Barnard 82. *(Photos, Mount Wilson Observatory, 100-inch reflector)*

expelled by the explosion have continued to expand in every direction for thousands or tens of thousands of years, and constitute today's nebula. We have, besides, actually witnessed the formation of tiny planetary nebulae round Nova Persei 1901 and Nova Aquilae 1918. These, however, did not remain visible for long, and normal planetary nebulae are very much larger. Nevertheless, their continued expansion condemns them also to eventual disappearance into the interstellar void.

In this connection it has been noticed that the largest planetary nebulae are also, on the average, the faintest. Their linear sizes are of the order of 10,000 A.U. ($5 + 10^{11}$ to 10^{12} miles, or 0.1 to 0.2 lightyears), and their mean densities hardly exceed a few tens of thousands of atoms per cubic centimetre, which represents a vacuum some thousands of times more perfect than can be produced in our laboratories. Only the enormous extent of so rarefied a gas, and the presence in the vicinity of a star emitting intense ultra-violet radiation, enable them to be seen at all. When we speak of their dimesions, it must be realised that we are referring to the epochs at which the nebulae emitted the light reaching us today, and most of the planetary nebulae are very distant (10^3 to 10^4 lightyears). We shall see later on that their distribution on the star sphere shows a very strong galactic concentration, analogous to that of novae and, generally, of all very distant stars.

Thus the bright diffuse nebulae and the planetary nebulae both bear witness to the presence in space of rarefied gaseous masses which are excited to fluorescence by very high-temperature O, Bo or B1 stars in their vicinity.

Bright nebulae are also observed in the vicinity of cooler stars, principally those of late B type. But the spectra of these are entirely different, and merely reproduce that of the involved star — a continuous spectrum, crossed by dark absorption lines. This proves that the nebula is shining by reflected and not by emitted light, in the manner of an illuminated cloud of dust. The absence of bright emission lines does not, however, necessarily indicate that there are no isolated atoms of an extremely rarefied gas in the nebulae, since even if there were, the ultra-violet radiation of the relatively cool stars in the vicinity would be too weak to excite them to emit their characteristic radiations.

The comparison of photographs taken in blue and red light shows that whereas the nebulosity involving the principal stars, of the Pleiades (type B3 — B5) is of a bluish tint, the large nebula (diameter, 5 lightyears) near to Antares is clearly reddish.

Leaving out of account the effects of individual irregularities of structure — often very remarkable — the brightness of these nebulae whether gaseous or dusty, shining by emitted or reflected light) decreases rapidly with increasing distance of the involved star. Beyond a certain distance they are no longer visible, but everything indicates that this is only because the star can no longer illuminate them with sufficient intensity. In fact, unlike the planetary nebulae, there appears to be no offspring-parent relationship between the diffuse nebulae and their involved stars; their shapes do not in the least suggest that they have been formed from these stars, but rather that their visible extent is the result of a chance spatial relationship with them.

This conclusion implies that the whole of interstellar space may not be the perfect vacuum it was once thought to be — that it is interpenetrated by more or less formless, highly rarefied clouds of gas and dust, among and through which the stars move. These clouds only become visible in localities where a sufficiently bright star is near enough to illuminate them. Everywhere else they will remain dark and unseen.

Dark galactic nebulae. — In fact, however, although these clouds of gas and dust are themselves invisible, their presence can under certain circumstances be

detected. Despite their extremely low densities they will absorb and scatter the light of any stars that lie beyond them, and if their thickness is great enough they will appear as dark clouds obscuring these more distant stars. Now photographs of the Milky Way are dotted all over with dark patches of just this sort; they were first studied systematically early in this century by the German, Max Wolf, and the American, Barnard. They are clearly defined dark areas, often of intricate shape, which are devoid of stars; they look like actual holes in the teeming star clouds of the Milky Way, but a moment's reflection will show the inherent improbability of innumerable 'tunnels' of this sort, all of which are aligned upon the Sun. They must in reality be clouds of matter which absorb the light of the stars situated beyond them. The few stars that are seen in these dark areas are relatively near to us, and are projected against the more distant dark nebulae. Apart from these considerations, the reality of the dark clouds is clearly demonstrated in certain particular cases where they are projected against, or associated with, bright nebulosity, as the reader can see for himself by studying the photograph of the Horse-head nebula near to ζ Orions (p. 372): jutting out from a large starless area, it resembles a great dark cumulus cloud fringed with light as it passes in front of the Sun. This association of dark and bright nebulae is, in fact, quite general; bright stars situated in these large dark regions are almost invariably surrounded by a bright halo, and conversely, nearly all the bright nebulae are involved to a greater or smaller extent with dark patches or lanes. The intimate relationship between dark and bright nebulosity is illustrated in many of the accompanying photographs.

Thus it seems that the Galaxy is encumbered with vast masses of material, complicated in structure, dark or bright according to the stellar population in their neighbourhood, which absorb (wholly or partially) the light of the more distant stars. The whole of the brighter half of the Milky Way, from Cassiopeia through Cepheus, Cygnus, Aquila, Scutum, Ophiuchus, Sagittarius, Scorpio, Centaurus, Crux and Argo is particularly rich in these dark nebulae. The largest (i.e. in general, the nearest) are clearly visible to the naked eye. Of these the most notable is certainly the 'Coal Sack' in the Southern Cross, though the northern hemisphere has one almost as obvious in the constellation Lynx, near the borders of Cygnus and Cepheus, as well as numerous others in this region. The great Cygnus rift that splits the Milky Way itself is also caused by dark clouds extending in the plane of the Galaxy at no great distance from the Sun.

It will be appreciated, therefore, that the dark nebulae exert an important, perhaps a decisive, influence on the apparent distribution of the stars, so that the observed structure of the Milky Way is attributed very largely to

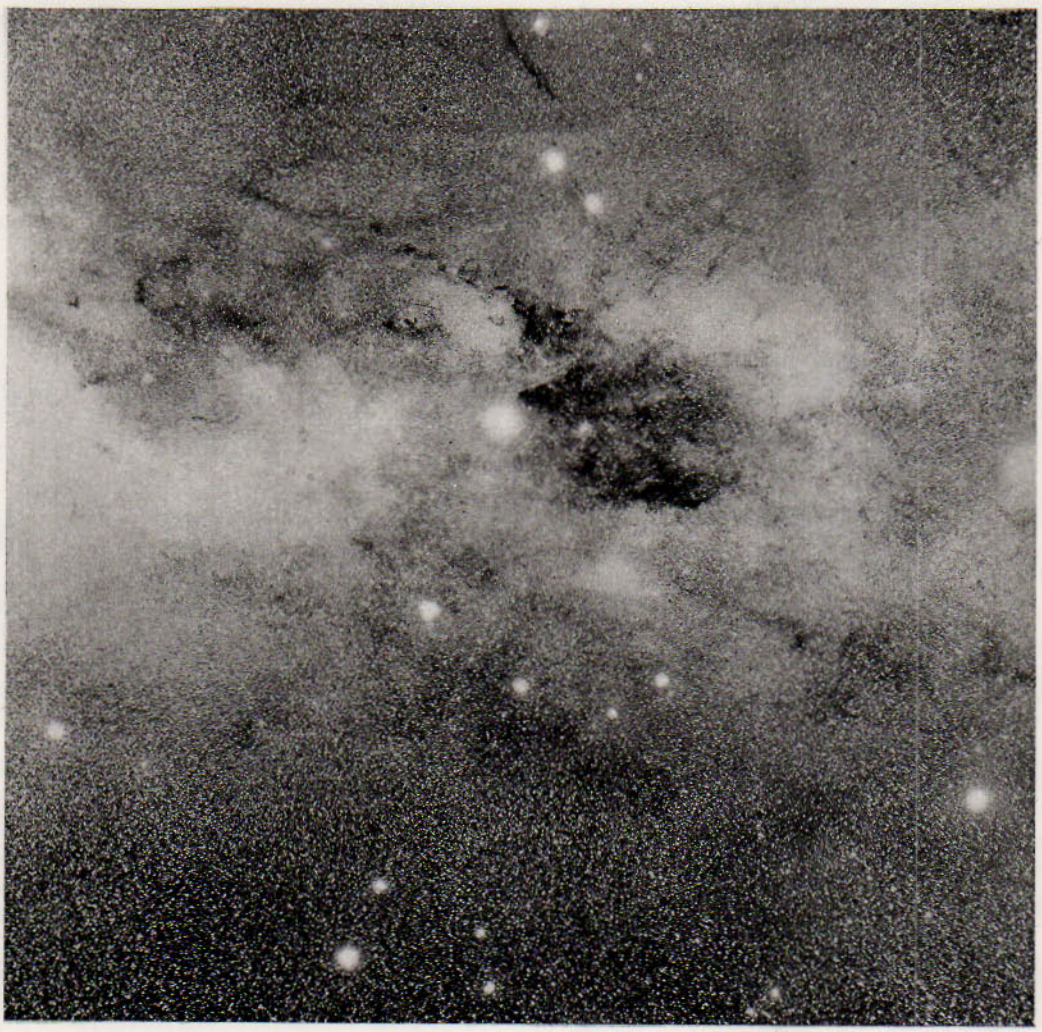


Fig. 644 — THE MILKY WAY IN THE REGION OF THE SOUTHERN CROSS, SHOWING THE COAL SACK.
(Photo, Boyden Station of Harvard College Observatory, South Africa)

Fig. 645 — THE HORSEHEAD NEBULA IN ORION, THE MOST SUGGESTIVE OF THE DARK NEBULAE. Projecting against a luminous background, and fringed with light, its resemblance to a dark cloud passing across the face of the Sun is very striking. *(Photo, Mount Wilson Observatory, 100-inch reflector, exposure 3h; 1920)*

them — in particular, the often sharply defined star clouds, which were for a long time considered to be real stellar clusterings. We shall return to this question later, when discussing the structure of the Galaxy.

To determine the extent of such an influence it is obvious that data must first be collected on the sizes and distances of the dark nebulae, as well as on the manner in which they reduce starlight passing through them. Max Wolf demonstrated that all these data can be obtained by

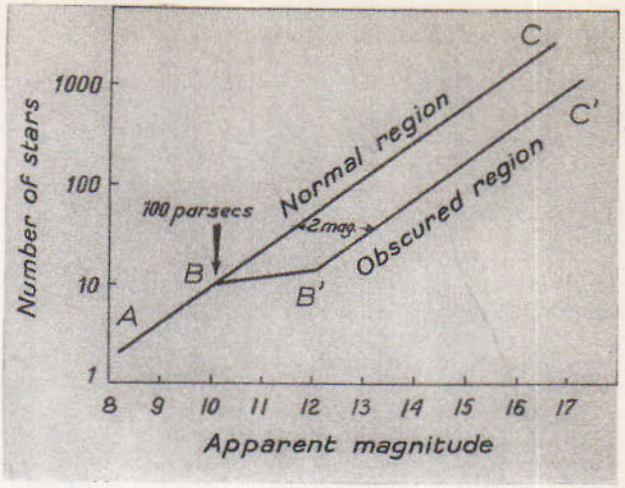


Fig. 646 — IDEALISED WOLF DIAGRAM OF A DARK NEBULA

means of comparative star counts in the obscured region and in nearby star fields that are not affected by obscuration. If a graph is constructed (Fig. 646) showing the star numbers in each of these regions as a function of magnitude, it is generally found that the two curves are more or less coincident down to a certain limiting magnitude; beyond this, the numbers in the unobscured region continue to increase normally, while the increase of numbers in the region of the obscuring cloud falls off considerably, since to a greater or lesser extent they exclude those stars lying within and beyond the cloud. As has been pointed out above, the wide range of stellar luminosities prevents us from reaching any exact conclusions if all the stars visible on the photograph are treated in this way. Stars of one spectroscopic type, corresponding to a single well-defined absolute magnitude, must be selected — Ao stars are valuable in this context, being both numerous and intrinsically brilliant.

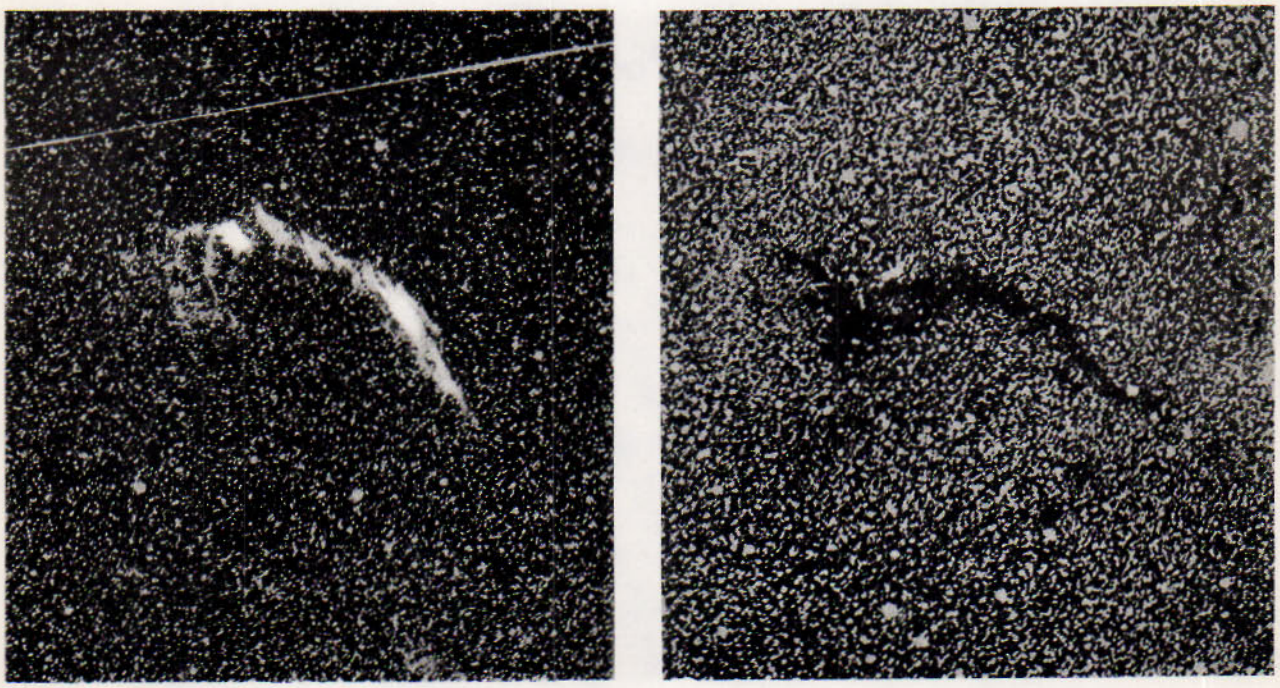


Fig. 647 — CURIOUS INSTANCE OF TWO DIFFUSE NEBULAE OF VERY MUCH THE SAME SHAPE, ONE OF THEM BRIGHT (NGC 6995 IN CYGNUS) AND THE OTHER DARK. *(Photos, E. E. Barnard, Yerkes Observatory, 10-inch Bruce telescope)*

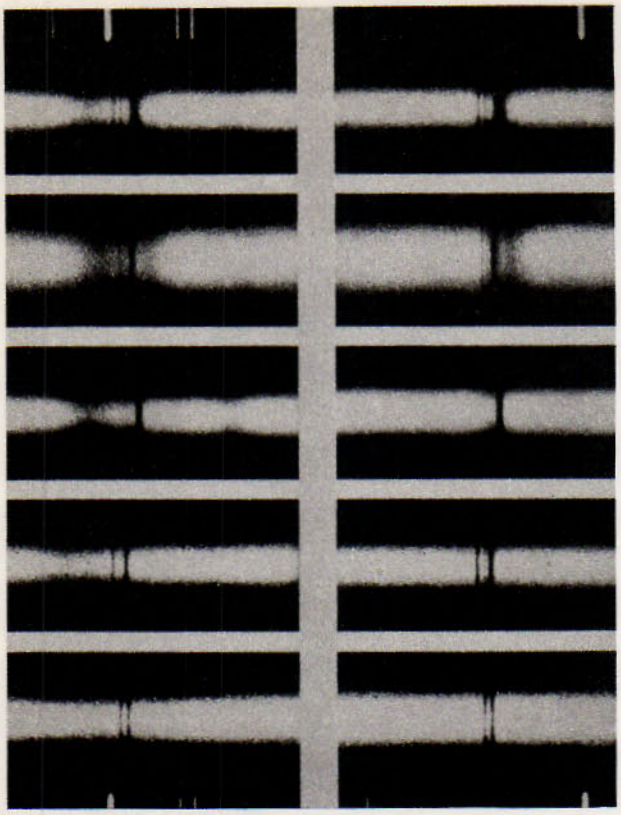


Fig. 648 — INTERSTELLAR LINES ARE DISTINGUISHABLE FROM THOSE OF STELLAR ORIGIN BY THEIR GREATER FINENESS. *(Mount Wilson Observatory)*

Fig. 646 shows the appearance of Wolf's diagram when this is done. At first, the number of stars encountered in both regions increases from magnitude to magnitude — i.e. with increasing distance. When the apparent magnitude corresponding to the distance of the nebula is reached, the numbers counted in the obscured region hardly increase any further for a while; however, they start to increase normally again as soon as the stars lying beyond the nebula are reached. The beginning of the near-horizontal section of the dark area curve thus tells us the distance of the cloud; its length, the total absorption of light from more distant stars, and its slope gives us some idea of the cloud's thickness.

In this way — and mainly during the past half-century — studies of the dark nebulae have revealed that their distances are for the most part of the order of several hundred lightyears, and their absorption 1 or 2 magnitudes. The Coal Sack, for example, is about 300 lightyears from the Sun, reduces the brightness of the stars beyond it by 60% to 70% (1 to 1.5 magnitudes), and has a total mass equal to at least 100 Suns.

Below are tabulated data of three other obscuring clouds that have been investigated in recent years:

CONSTELLATION	GALACTIC LONGITUDE	DISTANCE AND EXTENT		TOTAL ABSORPTION (PHOTOG. MAGS.)
		LIGHT YEARS	PARSECS	
Aquila	10°	420—880	130—270	1.1
		1300—5000	400—1500	about 4
Aquila	15°	1600	500	2
Ophiuchus	350°	6500	2000	
		650	200	>3.5

In addition to these large dark nebulae, certain much smaller and denser dark objects have engaged the attention of astronomers in recent years; they are approximately circular and very small, and are best seen when projected against a large area of bright nebulosity (Fig. 649). B. J. Bok, of Harvard College Observatory, has named them 'globules'. The mean dimensions of those occuring in the rich regions of the Milky Way seem to be of the order of 10,000 to 35,000 A. U. (1/6 to ½ lightyear), while those in the opposite regions of the sky are larger — 100,000 A.U. (or 1 to 2 lightyears). Their opacity is inversely proportional to their diameter, the absorption of the largest being at most 1 magnitude, while that of the smallest amounts to at least 2 — 5 magnitudes. This would appear to give them masses comparable with that of the Sun.

It has been suggested that they are clouds in process of condensation under the influence of the general radiation pressure of the star field all around them, and many astronomers believe that here, before our very eyes, are stars in the process of formation. The importance of this observation will be made more evident in Chapter 18.

In the foregoing account of the absorption effect of the dark nebulae, all the figures cited refer to the blue and violet wavelength range to which the photographic plate is most sensitive. These nebulae are in fact more opaque to the shorter wavelengths (blue, violet and ultra-violet) than to the longer wavelengths of yellow, red and infra-red light. Hence stars seen through an obscuring cloud appear to be redder than nearer stars of the same spectroscopic type, and their colour index is higher; for this reason they are said to show a colour excess. Now theoretical considerations indicate that the selective character of this absorption is connected with the dimensions of the particles exerting it, and the study of the reddening of light passing through the dark galactic nebulae has enabled the Swedish astronomer C. Schalen, among others, to derive the mean dimensions of their component particles. It appears that their diameters are about 0.1 microns, or 0.0001 millimetre. The material is therefore a very fine dust. However, this is only a mean value, and in fact particles of very different sizes are encountered from one nebula to another.

As to the chemical composition of the obscuring clouds, little has yet been established. They were originally supposed to consist of metallic dust (especially iron), whose existence in space is proved by meteorites. But very delicate observations initiated about 20 years ago by O. Struve and the Yerkes Observatory astronomers showed that the 'dark' nebulae are in fact very far from being perfectly dark, and that they appear so only by contrast with the brighter neighbouring regions. In reality, their reduction of the intensity of the starlight passing through them is much less a matter of absorption, in the strict sense, than of diffusion. This diffused light illuminates them, very faintly but nevertheless measurably, and the investigation of this weak luminosity seems to show that the dark nebulae diffuse more light than would be possible if they consisted of metallic particles. Their true nature, however, is still in doubt.

The density of the interstellar material varies considerably from one nebula to another, but it is estimated that in general it does not exceed 10^{-3} gm. per cubic kilometre*, and as often as not is 10 or 100 times less than this.

From the point of view of our everyday terrestrial experience, therefore, the expression 'interstellar void' remains perfectly apt. But the immensity of the spaces separating the stars is such as to compensate for even this degree of rarefaction, and on the galactic scale the diffuse

* One ton distributed through a volume equal to that of the Earth.

Fig. 649 — THE DIFFUSE NEBULA M 8 IN SAGITTARIUS IS DOTTED OVER WITH CURIOUS DARK NEBULOUS SPOTS, SMALL AND OFTEN SPHERICAL IN SHAPE. *(Photo, N. U. Mayall, Lick Observatory, 36-inch telescope; 12 July 1939)*

matter distributed throughout space does in fact play a very important part.

The general interstellar gas. — Interstellar material does not consist only of fine dust, constituting the dark nebulae. We have learnt that emission nebulae are to be seen around some of the hottest stars, betraying the existence of individual atoms of various gases, and there is no reason to suppose that these occur only in the immediate vicinities of such stars. Just as the nebulae shining by reflection are no more than the better illuminated regions of a vast cloud of cosmic dust, so the observed emission nebulae are simply privileged regions of great non-luminous clouds of rarefied gas. But the existence of this gaseous material is much more easily demonstrated than that of the dust cloud. It is true that the latter absorbs and diffuses the light of distant stars to a very noticeable degree — thus immediately revealing its existence — whereas the individual atoms of a rarefied gas exert an altogether inappreciable absorption, so that photographs of the Milky Way will give no indication of their presence. Nevertheless, their existence is revealed in a more subtle fashion — by the spectra of distant stars, on which they impress their characteristic absorption lines, just as do the atoms in the star's atmosphere. And furthermore, if the atoms producing these lines do not naturally occur in the star's atmosphere, they can be distinguished from the lines of stellar origin.

Thus, for instance, the two lines of calcium in the ultra-violet do not normally occur in spectra of the blue B type stars. In certain cases these lines are nevertheless to be observed, but both their behaviour and their appearance differ sharply from those of the other lines. They had indeed already attracted the attention of the German astronomer Hartmann in the early years of this century. Studying the Bo spectroscopic binary δ Orionis, he was astonished to observe that the K line of ionised calcium alone of all the lines in the star's spectrum did not oscillate back and forth about a mean position in step with the orbital motion of the pair (p. 314). Furthermore, this line — which he therefore characterised as 'stationary' — was thin and sharp in appearance, while the others, of hydrogen and helium, were wide and ill-defined.

Since then, stationary lines of ionised calcium — and also of sodium and other elements — have been detected in the spectra of numerous stars. All the same, it was thought at first that they were produced by gases in the immediate neighbourhood of these stars, and only within the last thirty years has their interstellar origin been recognised. That the absorption producing the stationary lines occurs all along the light path from the star to the Earth is proved by the fact that on the average the intensity of the lines is proportional to the star's distance. Their fineness is — as we shall see later (Chapter 20) — an indication of the very low temperature and extreme rarefaction of the interstellar gas.

The investigations of the last half-century (notably at Mount Wilson) have established the existence in interstellar space not only of various neutral and ionised atoms, such as those of calcium, sodium, potassium and titanium (all of which occur also in the stars), but of certain molecules — in particular, those of cyanogen (CN) and hydrides such as CH, NaH, etc. The interstellar gas seems to be extremely rich in hydrogen. If the characteristic absorption lines of this element are not to be seen, the reason is either that the radiations it can absorb lie in the ultra-violet (which is itself totally absorbed by the Earth's atmosphere), or that the ultra-violet radiation of the O and B type stars scattered through space has deprived if of its single electron, so that it is no longer capable of exercising absorption. To make up for this, however, if a hydrogen nucleus encounters a free electron, their recombination will be accompanied by the emission of observable radiation. Radiation of this type, deriving from the recombination of hydrogen, was actually detected shortly before the war by Struve and Elvey at the MacDonald Observatory in Texas: by means of a specially designed spectrograph they succeeded in photographing the characteristic emission spectrum of hydrogen (particularly that in the red) in various regions of the Milky Way. The radiations of ionised oxygen were also detected. It is clear that these weakly luminous regions, often several hundred lightyears across, are associated with groups of type O stars which are responsible for the ionisation of the gas.

Calculation based on these observations has confirmed that hydrogen is the most abundant constituent of the interstellar gas, as it is of the stars. For every calcium atom there are probably several score of potassium and sodium, perhaps several hundreds of oxygen, but certainly some millions of hydrogen.

It has also been estimated that despite the extraordinarily low density of this gas (100 times less than that even of the dark nebulae) it accounts for at least 90% of the total mass of interstellar material; compared with this, therefore, the combined mass of the much more noticeable dusty material is very insignificant.

It has been recognised for the past 30 years that interstellar dust is not confined to the dark nebulae, but is widely distributed throughout space. Indeed, since the interstellar gas is not restricted to the bright emission nebulae it would have been surprising if the dusty cosmic cloud had been found to be confined to the dark nebulae and those bright nebulae that shine by reflected light. How can the existence of cosmic dust in all directions throughout the Galaxy be demonstrated? It obviously cannot, like the dark nebulae, be detected by means of comparative star counts, since no region can be assumed *a priori* to be unaffected by its absorption. But the innumerable extragalactic nebulae (which we shall speak of in greater detail in the next chapter) can be used in place of the stars that are obscured by the cosmic cloud. These objects are other galaxies, or star systems, immensely distant, and external to our own; they appear to be more or less randomly distributed throughout all extragalactic space, which they populate in their teeming myriads. Now a curious fact was brought to light by counts of the extragalactic nebulae, a long time ago: although they are very numerous in those regions of the star sphere that are far from the Milky Way, their numbers fall off progressively as the Milky Way is approached, and within the Milky Way itself hardly a single one is to be seen. But it is incredible that these external galaxies, which are completely independent of our own, should really be distributed in a fashion dictated by the shape and structure of the Galaxy. On the other hand their apparent distribution can be simply accounted for by supposing that (as we have already been led to suspect) the plane of the Galaxy is occupied by a very rarefied cloud of absorbing dust, whose accumulation over immense distances obscures the still more remote object lying outside the Galaxy but in the projection of its plane.

Hubble's accurate studies at Mount Wilson of the variation in numbers of extragalactic nebulae as a function of galactic letitude (i.e. their angular distance from the Milky Way) have shown that this absorbing layer, in which the Sun is immersed, reduces the brightness of those extragalactic object situated at the galactic poles by 0.25 magnitudes (about 20%). From the poles to the galactic plane the absorption increases steadily, in proportion to the increasing thickness of the galactic cloud through which their rays have to pass — in exactly the

Fig. 650 — THE TRIFID NEBULA IN SAGITTARIUS (M 20). A magnificent example of the association of dark and bright nebulosity. *(Photo, Mount Wilson Observatory, 100-inch reflector, exposure 2h 20m; 1921)*

Fig. 651 — M 78 IN ORION AND ITS NEIGHBOURS ARE THE SMALL ILLUMINATED REGIONS OF A LARGE OBSCURING GLOUD. *(Photo, de Kérolyr, Forcalquier Observatory, 30-inch telescope, exposure 4h; 1933)*

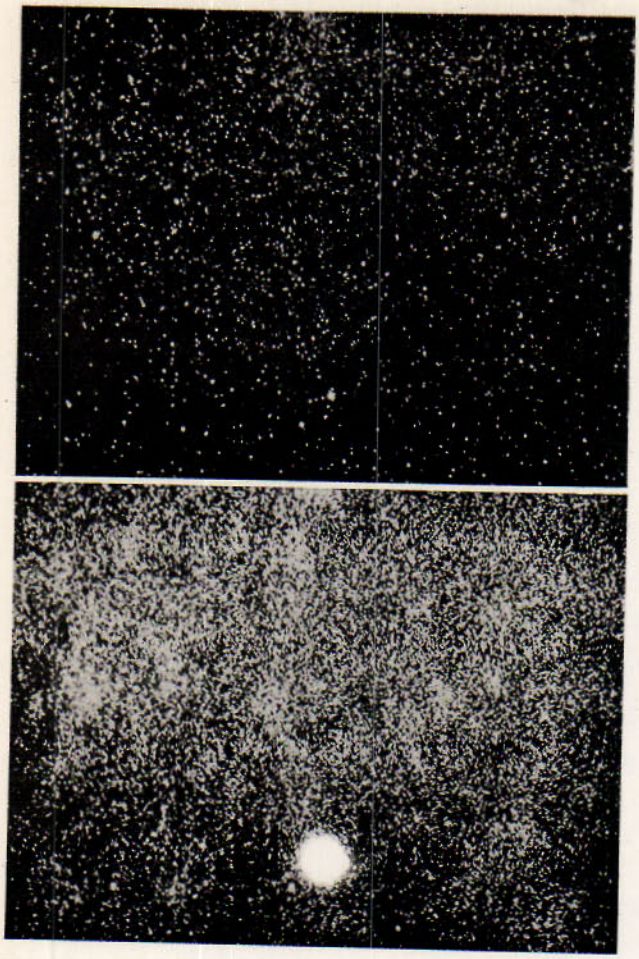


Fig. 652 — INTERSTELLAR REDDENING IN A SMALL AREA OF THE MILKY WAY. Two photographs taken by W. Baade with the Mount Wilson 100-inch reflector. Above: in blue light, ordinary photographic plate, exposure 60m; below: in red light, panchromatic plate and red filter, exposure 90m, which is equivalent to the exposure of the first photograph for type F7 stars (unreddened). Comparison of the two photographs shows that interstellar absorption is very much more marked in the blue than in red light, since the upper photograph bears no trace of the swarms of distant stars that crowd the lower.

same way the stars are progressively dimmed by our terrestrial atmosphere as they sink from the zenith towards the horizon (p. 305). After about galactic latitude 10°, practically no more extragalactic nebulae are to be seen. The width of this 'zone of avoidance' varies with galactic longitude, being greatest (about 40°) in the direction of Sagittarius, galactic longitude 330°. It is bordered on both sides by zones of partial avoidance, 10° to 20° wide on the average, though these also are wider in the Sagittarius region than elsewhere. Outside the two zones of partial avoidance the apparent distribution of the extragalactic nebulae becomes more or less even (see Fig. 667).

It is certain that within the zone of avoidance, and also to a large extent in the zones of partial avoidance on each side of it, the absorption of the light from extragalactic objects is due primarily to dark nebulae: piled up in depth, one behind the other, throughout the whole galactic plane, they are able to blot out these extragalactic objects completely. But the regular variation of their numbers outside the galactic zone shows quite clearly that absorption by relatively dense and localised clouds is superimposed on a weaker but more general absorption due to a more or less uniform and much more rarefied layer extending through the whole galactic plane like a veil of ultra-thin fog.

But if this fog extends all around us in the galactic plane, cannot evidence of its existence be found within the precincts of the Galaxy itself? This presents a difficult problem in that it involves the determination of the total absorption exercised by a given thickness of the interstellar cloud; nevertheless, as we shall see shortly, it is not impossible of solution. On the other hand it is comparatively easy to establish the selective character of this absorption — i.e. that it affects the violet and blue wavelengths more than the yellow and red — as has already been mentined in connection with the dark nebulae.

Because of this selective absorption, distant stars (whose light has passed through a greater thickness of cosmic material than that of nearer stars) appear to be redder than those of the same type situated in our vicinity. In much the same way we see the colour of the Sun's disc change from yellow to red as it sinks towards the horizon and its rays traverse a progressively greater thickness of atmosphere. Simple comparison of the colour indices of distant and near stars of the same spectroscopic type suffices to reveal the presence of the general cosmic cloud, because of the colour excesses shown by the more remote stars.

During the past twenty-five years measurements of this sort have been undertaken, notably at Mount Wilson by J. Stebbins and his colleagues, who have accurately determined the colour indices of over 1,300 O and B type stars photoelectrically. These stars are of great intrinsic luminosity (therefore visible over great distances) and are normally bluish, all of which facilitates the detection of any abnormal reddening. These investigations have also shown that such stars are strongly clustered about the galactic equator, and furthermore that they are especially numerous in the region from Perseus to Sagittarius (and probably beyond, in the more southerly region of the Milky Way from Centaurus to Argo, which is not observable from Mount Wilson). On the other hand, not a single one occurs between galactic longitudes 120° and 200°, in the region of the so-called 'anticentre', which is diametrically opposed to the galactic centre.

In this region, also, a proportion of the selective absorption can be attributed to the dark nebulae, but probably the greater part is caused by the general cloud. One begins to notice, moreover, that this is no longer regular and homogeneous; in certain areas, even quite close to the galactic plane, the reddening of distant stars is

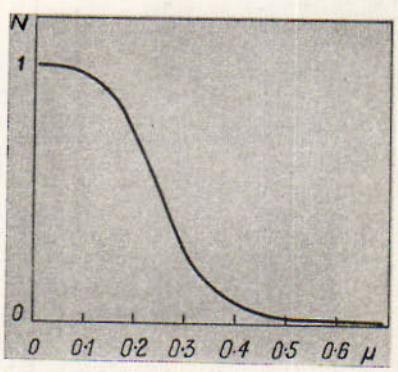


Fig. 653 — FREQUENCY OF INTERSTELLAR PARTICLES AS A FUNCTION OF THEIR DIMENSIONS (IN THOUSANDTHS OF A MILLIMETRE). *(J. H. Oort 1947)*

Fig. 654 — NEBULOSITY SURROUNDING ζ ORIONIS, INCLUDING THE HORSEHEAD NEBULA. *(Photo, de Kérolyr, Forcalquier Observatory, 30-inch telescope, exposure 11h; 1933)*

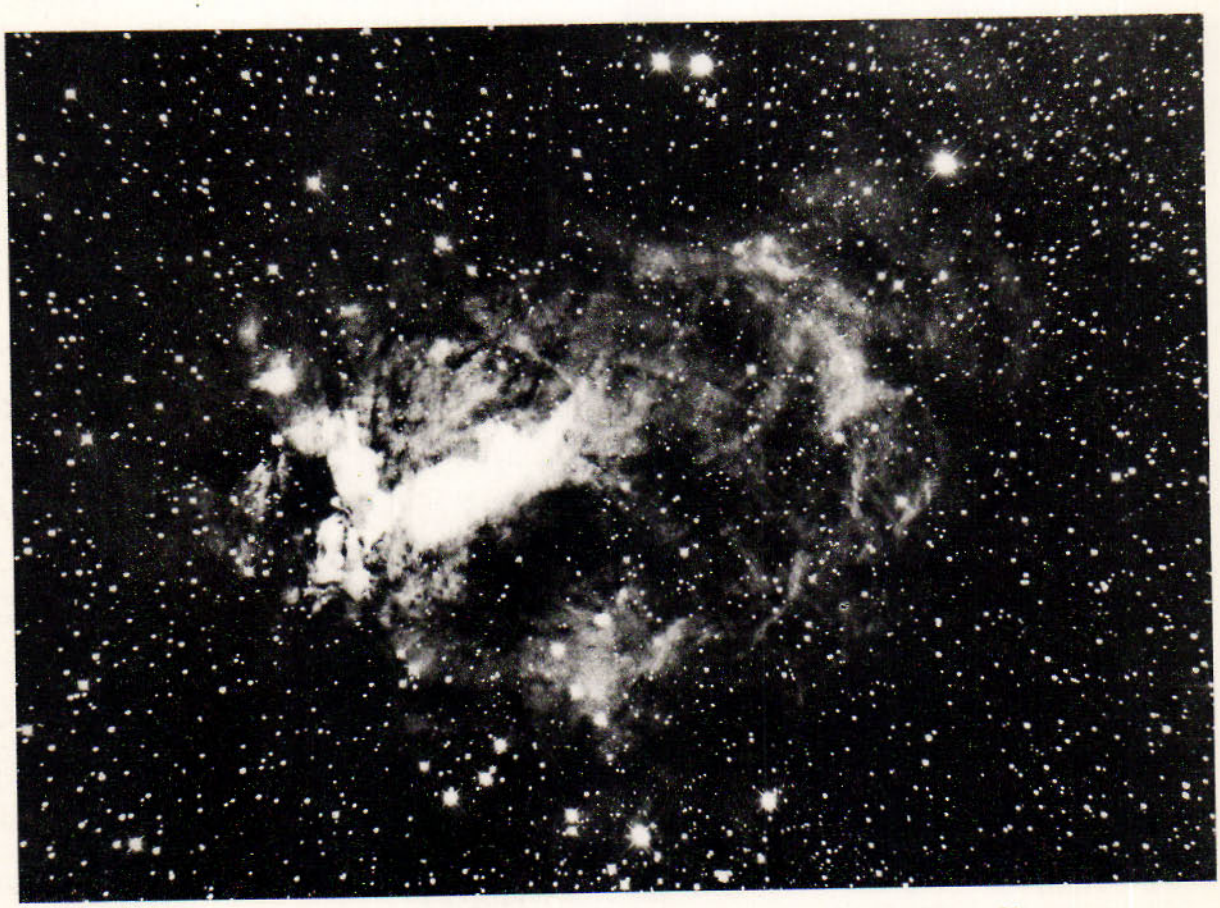


Fig. 655 — THE DIFFUSE OMEGA NEBULA, M 17, IN SAGITTARIUS. *(Photo, Mount Wilson Observatory, 60-inch reflector, exposure 3h; 1919)*

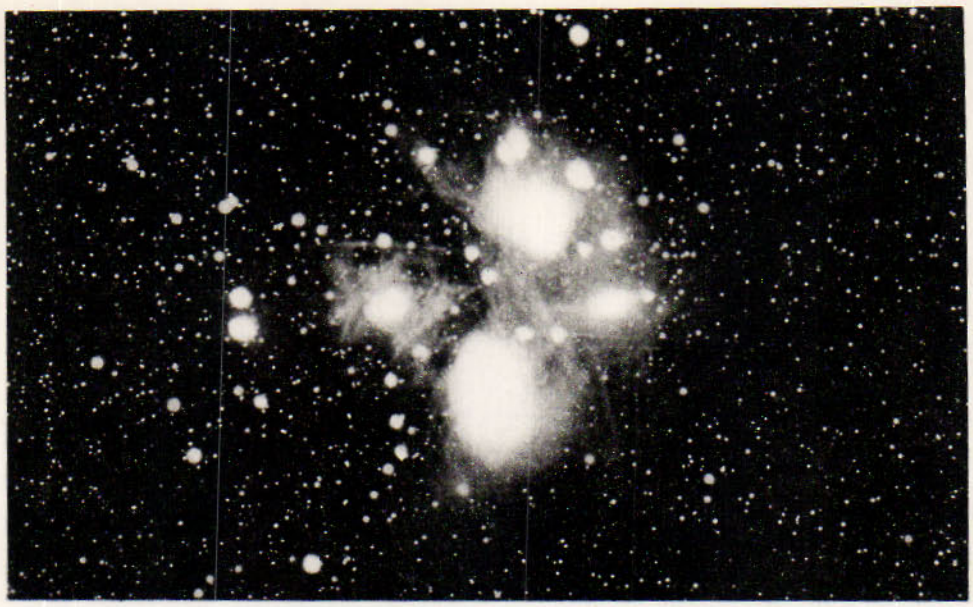


Fig. 656 — THE PLEIADES AND THEIR DUST CLOUDS. *(Photo, E. E. Barnard, Yerkes Observatory, 10-inch Bruce telescope)*

almost inappreciable. The law of stellar reddening also appears to vary from one region to another: for example, the spectrophotometric observations made a number of years ago by D. Barbier and D. Chalonge, of the Paris Institut d'Astrophysique, showed that the curves representing the varying absorption of different colours by the interstellar cloud were not the same for the two B type stars ζ Persei and χ Orionis, both of which are abnormally red.

All these investigations, which are still being energetically pursued, thus indicate that the stars are involved in a cloud of absorbing and diffusing matter distributed throughout the whole Galaxy — lacking definite structure, extremely rarefied, completely irregular, and composed of a wide variety of particles: free electrons, neutral and ionised atoms, molecules and solid particles ranging from ultra-microscopic grains of dust to (in all probability) meteorites of all sizes. The dust particles exert the most noticeable absorption effects, though the atoms (and especially those of hydrogen) make up the greater part by mass of the cosmic cloud. Its volume throughout the Galaxy is so immense that, despite its low density, calculation shows that it accounts for at least half of the mass of the Galaxy — which, as we shall see, is put at about 2×10^{11} times that of the Sun. There is thus as much matter diffused through space as there is condensed into stars; some astronomers are of the opinion that the proportion of diffuse to condensed matter is even higher. In Chapter 18 we shall discuss the cosmogonical implications of this conclusion, which was undreamed of even 40 years ago.

It will readily be understood that any investigation of the structure and dimensions of the Galaxy is considerably hindered by this ubiquitous cloud — unevenly distributed, here more rarefied, there more condensed, illuminated locally by giant stars, and almost everywhere limiting the depth of our visual penetration of space to a few hundreds or thousands of lightyears. To some

Fig. 657 — Left: the Pleiades as seen with binoculars; right: the physically associated members of the cluster are revealed by their parallel proper motions.

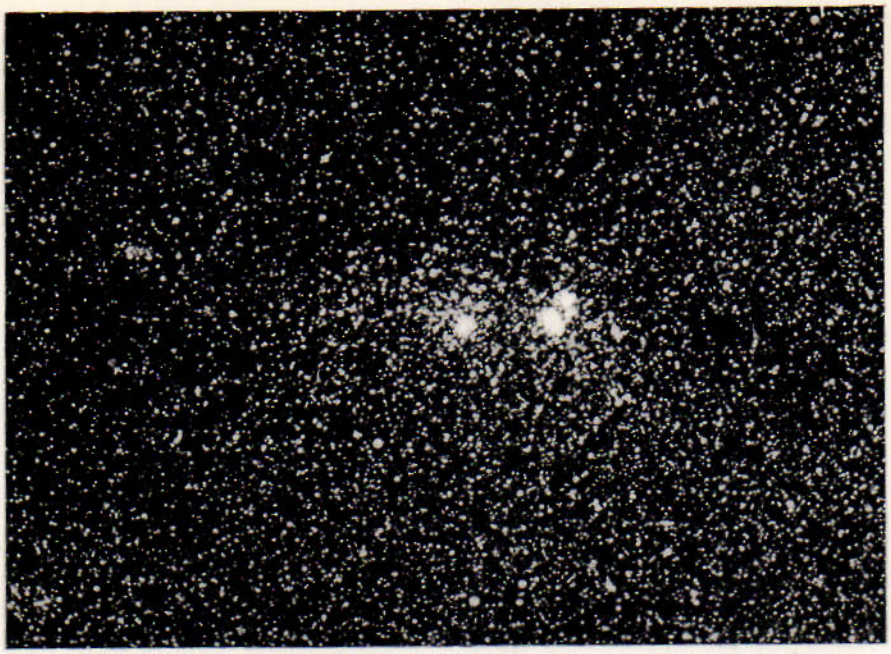


Fig. 658 — THE DOUBLE CLUSTER IN PERSEUS, SITUATED IN A RICH STAR FIELD OF THE MILKY WAY. *(Photo, E. E. Barnard, Yerkes Observatory, 10-inch Bruce telescope)*

extent these difficulties have been surmounted, however, thanks to information derived from the study of the star clusters; and to these objects we must now turn our attention.

Open or galactic clusters, and interstellar absorption. — We have already learnt that not all stars are solitary like the Sun, isolated from their neighbours in space: there are, for example, the numerous double, triple and multiple stars; then there are the much more widely separated groups whose physical connection is shown by their parallel proper motions and common radial velocity — we saw in Chapter 12 that 5 of the stars ot the Great Bear constitute a group of this kind. But here and there in the heavens we come across still richer clusters, comprising scores or hundreds of physically associated stars. The clusters with which we are concerned in this section are termed 'open', to distinguish them from the much more compact globular clusters which we shall consider next.

The largest — because the nearest — of the open clusters is the Hyades, well seen with the naked eye or a pair of wide-field glasses: it covers a large part of the constellation of Taurus in the neighbourhood of Aldebaran. The physical association of its stars is proved by their proper motions, which, by an effect of perspective analogous to that shown by meteors relative to their radiant, appear to converge upon a definite point on the star sphere.

The best known open cluster is certainly the Pleiades, situated not far from the Hyades on the borders of Taurus and Perseus; this beautiful little asterism is plainly visible to the naked eye, although a pair of field glasses shows it to better advantage. In passing, it may also be remarked that it provides an excellent test of visual acuity: a short-sighted person will see nothing but a blurred patch of light; normal eyesight generally reveals 6 or 7 stars, while the owners of excellent vision can detect about 10; children can often make out from 12 to 14. Long exposure photographs show at least 250 cluster stars brighter than magnitude 17, and the total number is probably about 500, not counting the foreground and background stars which are distinguishable from true members of the cluster by their proper motions. These 500 stars occupy an area of the star sphere some 6° in diameter, but the majority are confined to a central region measuring about 2° across; the latter also contains the brighter members that are visible to the naked eye. At its ascertained distance of about 500 lightyears the cluster's angular diameter is subtended by a linear distance of some 50 lightyears.

The star density within the Pleiades is at least 3 times that of the Sun's vicinity, for the number of stars within 16 lightyears (5 parsecs) of the Sun is only about 50 — or, more approximately, nearer to 30, since many of them belong to double or triple systems which would appear as single stars at the distance of the Pleiades. The density of the cluster's central region is nearer to 5 — 10 times that of the Sun's neighbourhood, so that the Pleiades does represent a very marked condensation in the general stellar field.

In varying degrees these characteristic are shared by all the open clusters; some are more condensed — the Double Cluster *h* and *χ* Persei, for example, which is visible to the naked eye as two faint patches of milky light in the galactic region between Cassiopeia and Perseus; others are less condensed, and therefore more difficult to differentiate from the crowd of superimposed stars of the foreground and background; some are richer — e. g. the Praesepe (Beehive) Cluster in Cancer (again clearly visible to the naked eye), and particularly some of the more distant ones such as M 11* in Scutum and M 37 in Auriga, which include several thousand members. Others, again, contain only a few score stars, if that; they can only be easily distinguished from accidental irregularities in the general field if they happen to be comparatively near to the Sun, as is the case with the stars of the Great Bear. It is, in fact, only among the large and relatively near clusters that the measurable proper motions of their members allow us to distinguish these from field stars; for this reason they are known as 'moving clusters'. All clusters are in motion, it is true, but the distances of the

* The number in Messier's catalogue of nebulae and clusters (see p. 400).

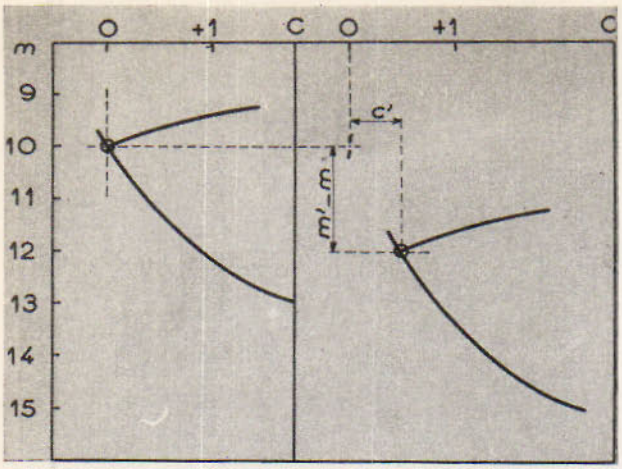


Fig. 659 — PRINCIPLE OF THE DETERMINATION OF INTERSTELLAR ABSORPTION BY MEANS OF COLOUR-MAGNITUDE DIAGRAMS OF OPEN CLUSTERS. *(J. Cuffey, 1939)*

others are too great for their proper motions to have become perceptible during the hundred years (and in many cases much less) that they have been accurately observed.

Up to distances of several thousand light years, nevertheless, open clusters can generally be recognised as such, and the number known within this range already runs into several hundred. They are strongly concentrated towards the plane of the Galaxy — hence their alternative name of galactic clusters — and for this reason they can give us valuable information about the characteristics of the Galaxy within a fairly wide radius from the Sun (see Fig. 668).

About 30 years ago R. J. Trumpler of Lick Observatory initiated the systematic study of the open clusters with this object in mind. For each cluster he constructed a diagram analogous to the Hertzsprung-Russell diagram, but with the latter's absolute magnitude, M, replaced by apparent magnitudes, m; m and M will differ from one another by a constant quantity for all stars of the same cluster.

Trumpler's diagrams revealed the existence of two main types of open cluster: in one, like the Pleiades, the Main Sequence from B to K is well represented, but the red giants are absent (Fig. 612); in the other, of which the Hyades is an example, the red giants are plentiful but the blue, B type giants are absent. The reason for this differentiation is not yet known, but the important result emerged from Trumpler's investigations of the Pleiades, Hyades and other nearby clusters of known distance that the absolute magnitudes of different types of star is pretty much the same in the open clusters as in the general field.

With this fact established, the distance of the cluster may be derived by determining the spectroscopic type (or, failing that, the colour index) of a certain number of stars in a cluster and comparing their observed apparent magnitudes with their absolute magnitudes as given by the Hertzsprung-Russell diagram. Suppose, for instance, that a certain cluster is found to contain some A type stars of apparent magnitude 10.5; their mean absolute magnitude, from the Russell diagram, is about +0.5, or 10^4 times brighter; their distance is therefore 10^2 times that of the standard distance which serves to define absolute magnitude (namely 10 parsecs) whence the distance of the cluster is 1 kiloparsec, or about 3,200 lightyears.

If this method is applied without any refinements it will, however, yield distances that are too great, since the light reaching us from remote stars is weakened by its passage through the interstellar cloud. It was precisely by means of his studies of the open clusters that Trumpler was in 1930 able to produce the earliest clear evidence of the existence of the general cosmic cloud. He had, indeed, established a fairly general relation between the total population and the degree of concentration of the nearest clusters, on one hand, and the mean linear diameter of the cluster, on the other; he could thus determine the distances of remote clusters from their apparent diameters, assuming that these were the same as for nearby clusters of the same type. Now though the distances derived by this method were in good agreement with those deduced from the spectrum-magnitude diagrams in the case of the nearer clusters, they were systematically too small for the more distant clusters. Since the inverse variation of apparent diameter with distance is a geometrical law that nothing can falsify, the conclusion was unavoidable that the distances given by the other method were too large — that, in fact, space is not perfectly transparent. The size of the discrepancy between the two sets of measurements allowed the coefficient of absorption of the interstellar material to be calculated, and Trumpler found a value of 0.7 magnitudes (photographic) per kiloparsec: in other words, the blue-violet component of a star's light that affects the photographic emulsion most strongly is reduced by about 50% every kiloparsec (3,200 lightyears).

The size of the error is thus a function of distance: the absorption for a cluster 5 kiloparsecs away (16,000 lightyears) will amount to about 3.5 magnitudes, and its stars will appear 25 times fainter than they would in absolutely transparent space; hence they are judged to be 5 times more distant than they are in reality — i.e. 25 kiloparsecs instead of 5. It is evident from this that interstellar absorption is an important factor in the determination of large distances by photometric methods; it is equally easy to understand that since Trumpler's discovery, the evaluation of this absorption with all possible accuracy has been a matter of great concern to astronomers. Recent estimates range from about 0.5 to 1.0 magnitude per kiloparsec, which on the whole confirm Trumpler's original figure.

We have seen that the precise determination of the absolute value of the general absorption is a difficult matter, but that the selective absorption producing interstellar reddening is much easier to measure. The possibility therefore suggested itself of establishing a relation between the two quantitites, by means of which the total absorption could be derived from data relating to selective absorption — i.e. from the observed colour excess of distant stars.

One of the most effective methods is that which was extensively used in America by Cuffey at the Goethe Link Observatory, some 20 years ago. It, also, involved the photometric analysis of the open clusters by means of Trumpler diagrams, but replaced the latter's spectroscopic types (which are difficult, or even impossible, to determine when the stars are very faint, as in the case of very distant clusters) by colour indices, which are easy to measure and which, as we have seen, are closely related to spectroscopic type. Now these diagrams, like that of Hertzsprung and Russell, contain one feature that is especially easy to recognise and define — the point at which the red giant branch joins the Main Sequence. It is located in the neighbourhood of absolute magnitude 0, spectroscopic type A, and colour index 0. Suppose, for example, that such a star is observed in a remote cluster (whose distance, derived from its apparent diameter, is known); then its distance modulus is $m - M = m$ (since $M = 0$), and we know (see Fig. 659) that this junction point of Main Sequence and giant branch lies at magnitude m' and colour index c', i. e. below and to the right

Fig. 660 – THE GLOBULAR CLUSTER ω CENTAURI, PHOTOGRAPHED AT THE SOUTHERN STATION OF HARWARD COLLEGE OBSERVATORY. Note its slightly elliptical shape.

of the position it ought to occupy; from this we can at once deduce the total absorption $m' - m$ and the colour excess c', which is a function of the selective absorption.

The investigation of many clusters along these lines (as well as others which we cannot describe here) has shown that on the average the relation of total absorption in blue light to selective absorption between the yellow and the blue is about 4:1. In other words, if a very distant star has an apparent photographic magnitude of 14 and is at the same time abnormally red for its spectroscopic type (having a colour excess of 0.5 magnitude, for instance), its light is undergoing a total absorption of about 2 magnitudes, and it is really of magnitude 12 – hence it is 2 ½ times nearer than its brightness would lead one to suppose.

Recent work has in fact shown that both interstellar reddening and absorption vary widely from one region of the sky to another. It is estimated that in the most transparent regions of the Milky Way the coefficient of general interstellar absorption varies from at least 1 magnitude per kiloparsec in the direction of the galactic centre (Sagittarius) to 0.5 magnitude per kiloparsec in the direction of the anticentre.

Nevertheless, the mean values just quoted have proved to be extremely useful when applied to large collections of objects situated in all parts of the sky, such as the open clusters. They have shown that the most remote open clusters, reduced to minute angular dimensions and enfeebled by interstellar absorption, finally sink below the limit of vision at distances exceeding 3 to 4 kiloparsecs (10,000 to 13,000 lightyears). Great though such distances are, they are still only a small fraction of the overall dimensions of the Galaxy. In order to make even deeper soundings into galactic space, appeal must be made to clusters which are more easily identifiable at great distances than the open clusters: with their help we shall finally succeed in measuring the diameter of the whole star system.

Globular clusters, and the size of the Galaxy. – The clusters referred to in the last few lines are enormous stellar conglomerations, approximately spherical in shape, and containing several 10^5, probably 10^6, members, as against the open clusters' hundreds or at most thousands.

Three globular clusters are visible to the naked eye: α Centauri and 47 Tucanae in the southern hemisphere, which look like small, rather nebulous stars of the fourth magnitude, and M 13 in Hercules, which is visible in our northern skies as an ill defined magnitude 5 star.

At present about 100 globular clusters are known, and it seems unlikely that their total number much exceeds this – or, at any rate, that many more will be discovered in the future. For most of them were already known to Herschel at the end of the 18th century, very few having been discovered since then despite the introduction of photography and the employment of telescopes far larger than his. This fact alone suggests that the globular clusters are very distant objects of high luminosity, and are not uniformly distributed throughout all space, for otherwise their numbers would increase in step with the power of the instruments used to observe them.

The connection between the globular clusters and the Galaxy is clearly indicated by their distribution on the star sphere: they show very strong galactic concentration, particularly in the direction of Sagittarius, where nearly one-third of their total number are to be found. Moreover, careful examination reveals that they are concentrated along either side of the Milky Way rather than on the galactic equator itself, where none is to be seen. This distribution is at once explicable if we assume that the globular clusters are exceedingly distant objects, for those lying in the galactic plane would then be hidden by the dark clouds that are piled up in this region, just as the similarly placed extragalactic nebulae are hidden.

It would therefore seem that the globular clusters – intrinsically very brilliant and easy to recognise, at the same time very remote and distributed in a manner which

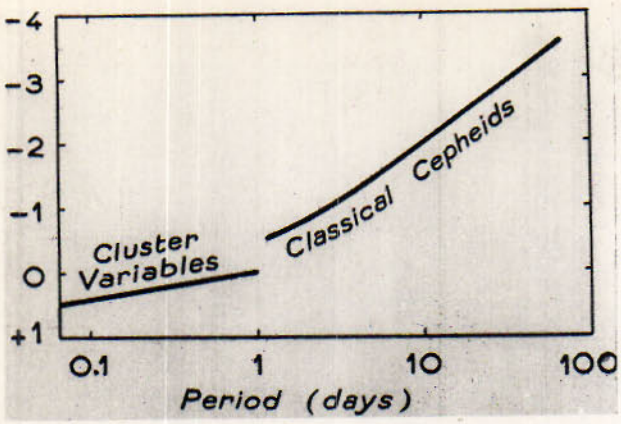


Fig. 661 — THE PERIOD-LUMINOSITY RELATION FOR CLUSTER VARIABLES AND CLASSICAL CEPHEIDS, AS EMPLOYED BY SHAPLEY

is intimately related to the structure of the Galaxy — are the ideal objects for fixing the overall dimensions and shape of the stellar system. Soundings based upon them carry us to much greater distances than those of the open clusters: it has in fact been found that the distance of the nearest globular clusters is slightly less than that of the most remote open clusters.

The first astronomer to recognise the value of the globular clusters in this connection was H. Shapley, former Director of Harvard College Observatory, who between 1915 and 1920 employed the great reflectors at Mount Wilson to make a systematic survey of them. This investigation proved that the globular clusters — though varying among themselves in richness and degree of condensation — form a homogeneous class which is altogether distinct from the open clusters. The majority are virtually spherical, though some (e. g. ω Centauri) are slightly elliptical. Colour-magnitude diagrams, analogous to those used in the study of the open clusters, clearly showed the presence of a well defined giant branch: the brightest stars in the globular clusters are red giants, several hundreds of times more luminous than the Sun. Since their apparent magnitudes are frequently as low as 10, 12, or even 14, it was at once obvious that the distances of the globular clusters are to be measured in tens of thousands of lightyears. Nevertheless, the population of these clusters is much richer in dwarfs than giants, though they may be too faint and too closely crowded to be studied individually; hence the integrated spectrum of a typical globular cluster is more often than not of the solar type. A few clusters, however, belong to types F and K, and exceptionally even to A or M.

But Shapley's most interesting results, from the point of view of the relation of the system of the globular clusters to the Galaxy as a whole, were provided by his systematic study of the variables in these clusters. Although we can still point to hardly a single intrinsic (i.e. non-eclipsing) variable which is indubitably a member of an open cluster, the presence of numerous variables in the globular clusters had been recognised by different observers since the beginning of the 19th century. S. I. Bailey, in particular, had undertaken a thorough survey of them in about 1900, using the large collection of photographs that had been taken at Harvard College Observatory and its southern station. He discovered that the majority of them were short-period Cepheids of the RR Lyrae type — for which reason they are often called 'cluster variables', although they also occur outside globular clusters. More than 100 variables can often be counted in a single cluster (M 3 in Canes Venatici and ω Centauri are examples), while in others they are scarce or even absent (e. g. M 13 in Hercules, or M 56 in Lyra). The reason for these differences is not known, but, whatever it may be, the fact of fundamental importance is that these variables obeyed the period-luminosity relation — discovered by Miss Leavitt (see p. 321) — whose general validity had been accepted by Shapley several years before he undertook his study of the globular clusters. Thus he was immediately able to determine the distances of those clusters containing variables whose periods had been measured. In M 3, for example, the mean photographic magnitude of the cluster variables is 15.5; since their absolute magnitude is 0.0 the distance was at once derived as 12 kiloparsecs, or about 40,000 lightyears. Since the time of Shapley's pioneer work the distances of many globular clusters have been determined in this way.

Sad to relate, however, the results obtained by Shapley with so much ingenuity have had to be extensively revised in the light of the recent discovery that the Cepheid class is in fact divided into two quite distinct categories — cluster variables and classical Cepheids. The effect which this discovery has had in the measurement of astronomical distances is given in detail in Chapter 13. It is interesting, and at the same time serves as an object lesson in the care that astronomers must exercise in interpreting their data, to learn that Shapley had what might have been the key to this fact long before it was actually announced in 1952.

The circumstances were as follows: Shapley, extending his principles further, worked out indirect methods for determining the distances of the globular clusters. The variables are far from being the brightest stars in the clusters — the red giants and supergiants are quite 5 or 6 times brighter — and once Shapley had determined by means of the then accepted period-luminosity relationship the magnitude of the Cepheids in the clusters, he discovered that there was a discrepancy of from 1.0 to 1.5 magnitudes between the distances of the globular clusters as determined from observations of their Cepheids and of their giants and supergiants. Unfortunately, he took it for granted that the Cepheid curve as then known was inviolable, and concluded that the red giants and supergiants in the clusters must be 1.5 magnitudes fainter than the value accepted as normal for these stars elsewhere. This assumption is the more surprising in view of the fact that Fig. 661 (which shows the values obtained by Shapley) is remarkably similar, with its discontinuity, to the currently accepted general curve shown in Fig. 581. It is now well established that the integrated absolute magnitudes of the globular clusters mostly lie between -5 and -7, with a mean value of -6. Those of some exceptional clusters (such as ω Centauri and 47 Tucanae) are as high as -8 or even -9, corresponding to several hundred thousand times the luminosity of the Sun.

The linear diameters of the globular clusters are more difficult to define, because although the stars are extremely numerous in their central regions (where their individual images combine into an undifferentiated blur of light on long exposure photographs) they thin out progressively towards the borders of the cluster, to which no sharply defined limit can be assigned. Nevertheless it can be said that the mean diameter of the more visible part of a typical globular cluster is of the order of several tens of parsecs — 100 lightyears or more. The diameter of the area within which widely scattered members of the cluster are to be found may be put at about 100 parsecs, or 300 lightyears.

All that can be said of the star density within a globular cluster is that it is extraordinarily high: whereas the distance between neighbouring stars in the Sun's vicinity is

Fig. 662 — THE GLOBULAR CLUSTER M 13 IN HERCULES, PHOTOGRAPHED BY G. W. RITCHEY WITH THE MOUNT WILSON 60-INCH REFLECTOR, TOTAL EXPOSURE 11 HOURS, SPREAD OVER, 6, 7, 8 JUNE 1910. 40,520 individual stars have been counted on this plate.

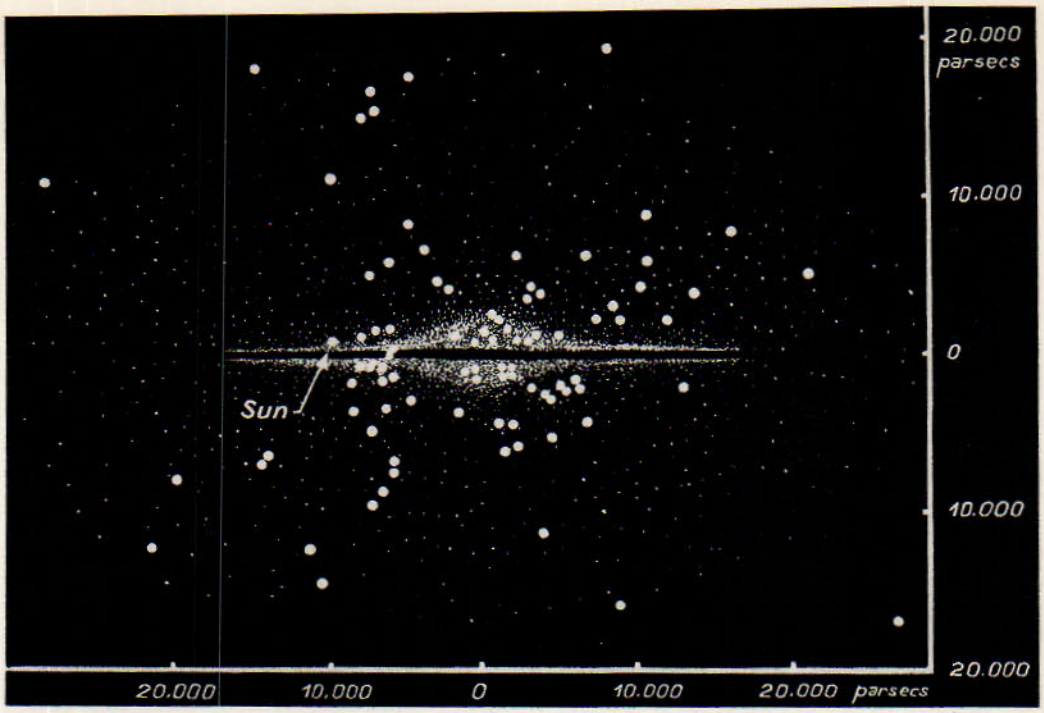


Fig. 663 — DISTRIBUTION OF THE GLOBULAR CLUSTERS IN RELATION TO THE PLANE OF THE GALAXY.
(J. S. Plaskett, 1935)

several lightyears, and within the most condensed open clusters about 1 lightyear, it must be of the order of several lightdays, or at most lightweeks, in the globular clusters — perhaps only a matter of lighthours in their most condensed central regions. The night sky of any planets that may be revolving about these stars must be of unexampled brilliance, crowded with stars brighter than the full Moon.

The most important result of Shapley's work on the globular clusters, however, was that it gave first reliable indication of the size of the Galaxy — assuming, that is, that the system of the globular clusters and the Galaxy are co-extensive. The distances of the clusters being known, it was a simple matter to determine their spatial distribution; when this was plotted it was found that they form a relatively unflattened system whose maximum extension in the galactic plane appeared to be nearly 80 kiloparsecs (250,000 lightyears), and whose centre lay at a distance of 50,000 lightyears from the Sun in the direction of Sagittarius. It had once been thought that the Sun was situated near the centre of the star system, and that the latter's diameter did not exceed about 30,000 lightyears. Shapley thus dethroned the Sun from its privileged position, just as Copernicus had dethroned the Earth four centuries earlier.

In fact, however, Shapley's original estimates were excessive: the considerable part played by interstellar absorption had not been discovered at that time, and his distances for the clusters situated in the Milky Way were accordingly too large. Modern estimates, taking interstellar absorption into account, are about 30 kiloparsecs (100,000 lightyears) for the diameter of the Galaxy, and 10 kiloparsecs (30,000 lightyears) for the distance of the Sun from its centre.

This naturally leads us to the question of the structure of the Galaxy, as it is conceived today.

The general and local structure of the Galaxy. — In this chapter we have already discussed some of the reasons for believing that the Sun occupies an eccentric position within the Galaxy; and, in so far as this conclusion was finally put beyond doubt by the study of the globular clusters it is of quite recent date.

It may therefore be worth while reviewing the principal reasons for believing that the Sun lies far from the centre of the Galaxy and that the centre lies in the direction of the constellation Sagittarius. Some of these we have met already. They are:

(1) The naked-eye appearance of the Milky Way — brighter and wider in the hemisphere from Cygnus to Carina, and especially so in the constellations Ophiuchus, Sagittarius and Scorpio (cf. Fig. 665).

(2) Star counts, which show that the richest and most extensive star fields lie in this same region.

(3) The concentration, again in the same region, of the most luminous and distant objects, such as novae and the planetary nebulae that are probably derived from them. This concentration is most marked in the case of the faintest objects, which may be assumed to be also the most distant. (See p. 388.)

(4) The concentration of diffuse nebulae, both bright and dark, in the same regions, and their comparative rarity in the direction of the galactic poles.

(5) The absence of O and B type stars with a large colour excess from the latter regions, as compared with their abundance all round the galactic equator, and particularly in the Sagittarius region.

(6) The distribution of the open clusters, which are both less numerous and apparently less condensed (since less distant) towards the galactic plane in the region from Auriga to Canis Major, whereas they show strong galactic concentration in the Aquila-Sagittarius-Crux region. (See p. 388.)

(7) The distribution of the globular clusters which because of their great distances, are absent from the galactic equator itself but are strongly concentrated on either side of it; furthermore, they are particularly numerous in Sagittarius and Scorpio, while only one or two are to be observed in the opposite half of the star sphere from Cygnus to Canis Major via Auriga.

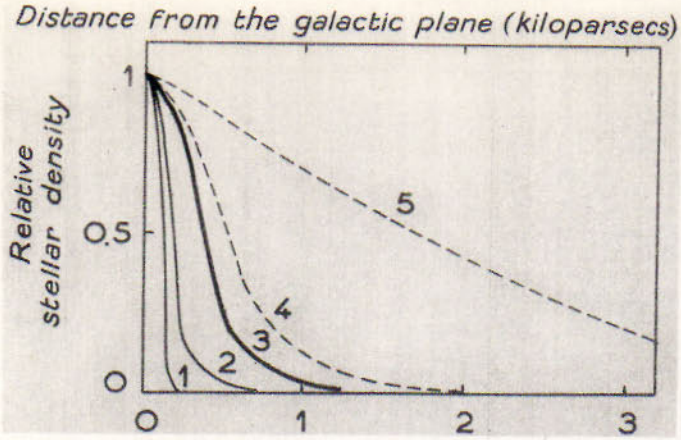


Fig. 664 — THICKNESS OF THE MILKY WAY IN THE VICINITY OF THE SUN. 1, type B stars; 2, classical Cepheids, open clusters and type A stars; 3, type F stars; 4, type K giants; 5, short-period Cepheids and long-period variables.

(8) The width of the zone of avoidance of the extragalactic nebulae: maximal in the Scorpio-Sagitarrius region, minimal in the opposite direction of the star sphere, where a few nebulae can even be observed within this zone, through 'windows' opening on to outer space. (See p. 388.)

(9) Finally, we shall see in this section that the motions of the stars have revealed the whole Galaxy to be in rotation about an axis which, once more, lies in the direction of Sagittarius at a considerable distance from the Sun.

This mass of concordant observations leaves no doubt that the Sun is a member of an immense, flattened congeries of stars, strongly condensed both towards its median plane and towards its centre, and that the Sun lies near this plane but at a considerable distance from the centre of the system.

The diameter of the Galaxy appears to be in the region of 80,000 to 100,000 lightyears, and its centre to lie in the direction of Sagittarius (more precisely, in galactic longitude 325°, on the borders of Sagittarius, Scorpio and Ophiuchus) at a distance of from 25,000 to 30,000 lightyears.

Estimates of the thickness of the Galaxy in the neighbourhood of the Sun differ according to how it is defined, since different categories of celestial object are not equally concentrated upon the galactic plane. Some — like the globular clusters, short-period Cepheids and type M long-period variables — exhibit only a small degree of galactic concentration, and are to be encountered up to distances of several thousand parsecs from the galactic plane. But these objects are of relatively rare occurrence, and the vast majority of the stars cluster about the median plane to a thickness not exceeding several hundreds of parsecs: 100 to 300 parsecs for stars of types A and B, open clusters and classical Cepheids, and 1,000 to 1,500 parsecs for giants of types G and K. The interstellar cloud also seems to be strongly concentrated towards the galactic plane: its total absorption of 0.5 magnitudes as measured by Hubble, and Trumpler's coefficient of absorption of 0.7 magnitudes per kiloparsec, indicate a total thickness of about 700 parsecs, similar to that of the stars.

The thickness of the Galaxy in our vicinity hardly exceeds a few thousand lightyears, or a few hundredths of its diameter; the term 'galactic *plane*' thus appears to be entirely apposite.

It has also been established that the Sun does not lie exactly in the galactic plane, but about 30 parsecs (100 lightyears) to the north of it, in the direction of the constellation Coma. This is indicated by the fact that the galactic equator is not quite a great circle, exactly equidistant from the poles, but is displaced slightly towards the south, as can be seen from a careful examination of the planispheres on pp. 387 — 389.

The thickness of the central region is much more difficult to determine. In view of the great distance of the galactic nucleus and the effect of interstellar absorption in this direction, it seems unlikely that the width of the Milky Way in Sagittarius represents the thickness of the nucleus itself. It clearly indicates, however, that the thickness of the stellar system does increase towards its centre: the region of the Milky Way that is richest in faint stars is about 20° wide, corresponding to 1 kiloparsec (2,000 — 3,000 lightyears) at a distance of 2 to 3 kiloparsecs, which seems to be very roughly the distance of the faint members of the star clouds bordering the central absorbing zone, although it is no more than a quarter of the way from the Sun to the nucleus. Further than this we cannot penetrate by visual or ordinary photographic methods of observation, but since the absorption affects the red rays less than the blue (and the infra-red less still) it was hoped that it might be possible to reach the nucleus by means of infra-red photometric measures of the Milky Way in the Sagittarius region. This seems actually to have been achieved by Stebbins and Whitford, who during 1945—6 at Mount Wilson undertook a series of photoelectric measurements in blue, red and infra-red. The latter showed an elliptical thickening of the Milky Way in longitude 326.5° which measures 8° × 3.5° and is coincident with the galactic equator; its reddening is such that they estimated that less than 0.1% of its light in ordinary photographic wavelengths reaches the Earth. Assuming this value, which corresponds to a total absorption of about 7 magnitudes, Trumpler's coefficient (0.7 magnitude per kiloparsec) gives its distance as about 10 kiloparsecs, which agrees exactly with the expected distance of the nucleus from the Sun. Its diameter would then be 1.5 and its thickness 0.6 kiloparsecs (5,000 and 2,000 lightyears, respectively). Such, at any rate, would seem to be the dimensions of the most easily seen region of the nucleus, and it is likely that its overall dimensions are considerably larger.

The contemporary picture of the Galaxy therefore envisages a very flattened system some 100,000 lightyears in diameter, consisting of a central nucleus with a diameter of 8,000 to 10,000 lightyears and a thickness of perhaps 3,000 to 4,000 lightyears, surrounded by a peripheral zone whose thickness decreases progressively from

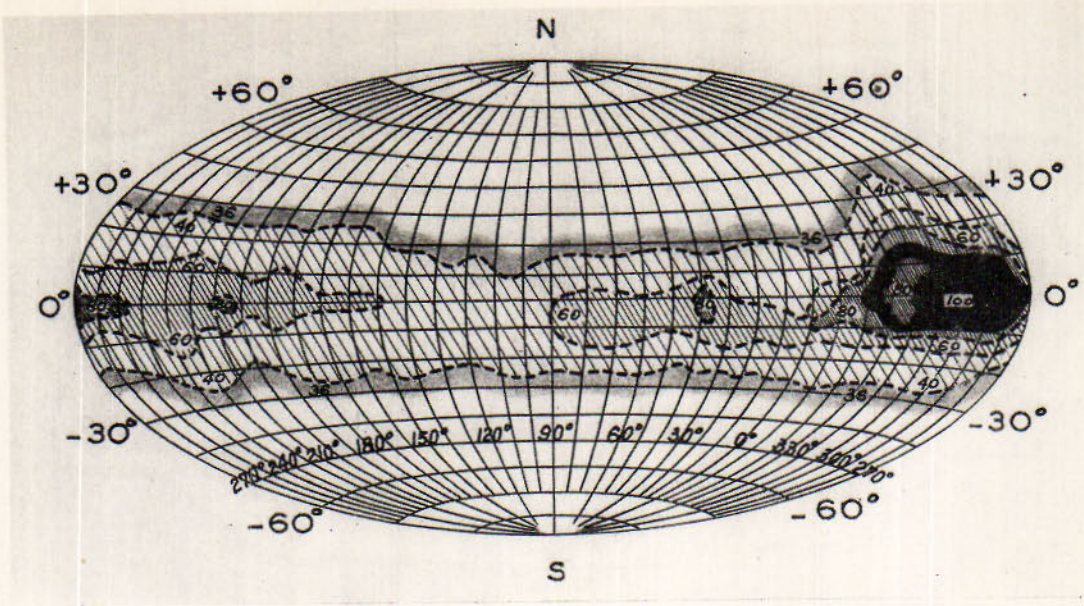


Fig. 665 — VISUAL ISOPHOTES OF THE MILKY WAY (HOPMANN, 1923)

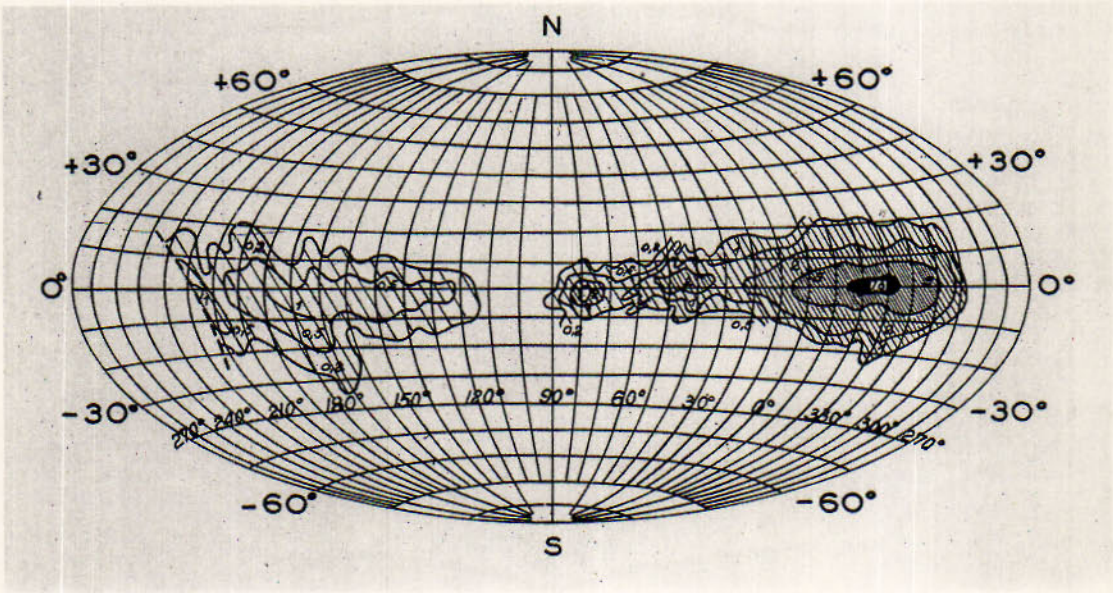


Fig. 666 — RADIO ISOPHOTES OF THE MILKY WAY FOR A WAVELENGTH OF 1.87 METRES (G. REBER, 1945)

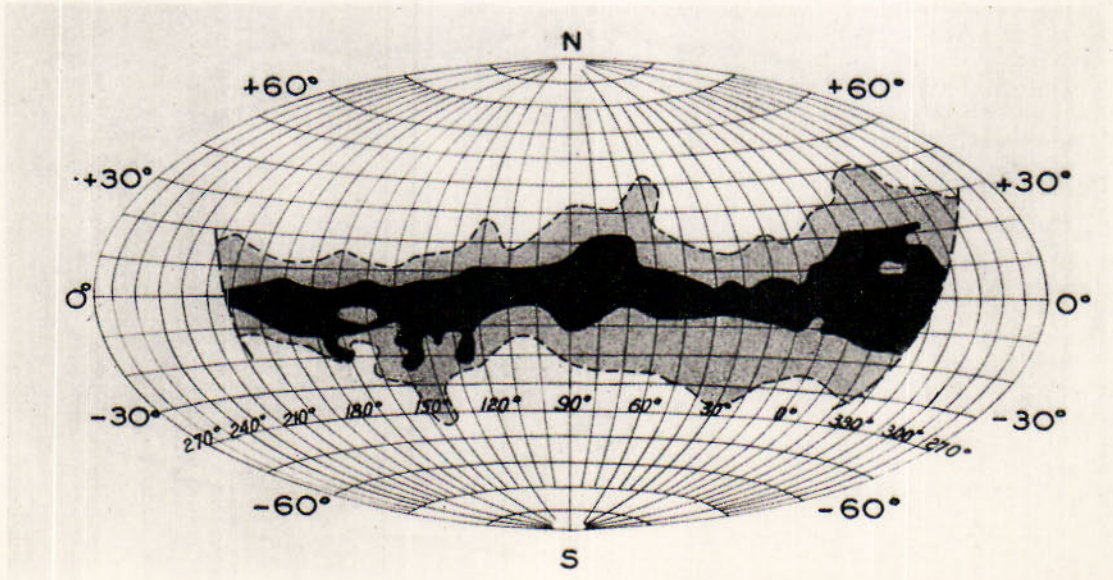


Fig. 667 — ZONES OF TOTAL AVOIDANCE (BLACK) AND PARTIAL AVOIDANCE (GREY) OF THE EXTRAGALACTIC NEBULAE, SHOWING THE EFFECT OF INTERSTELLAR ABSORPTION (HUBBLE, 1935).

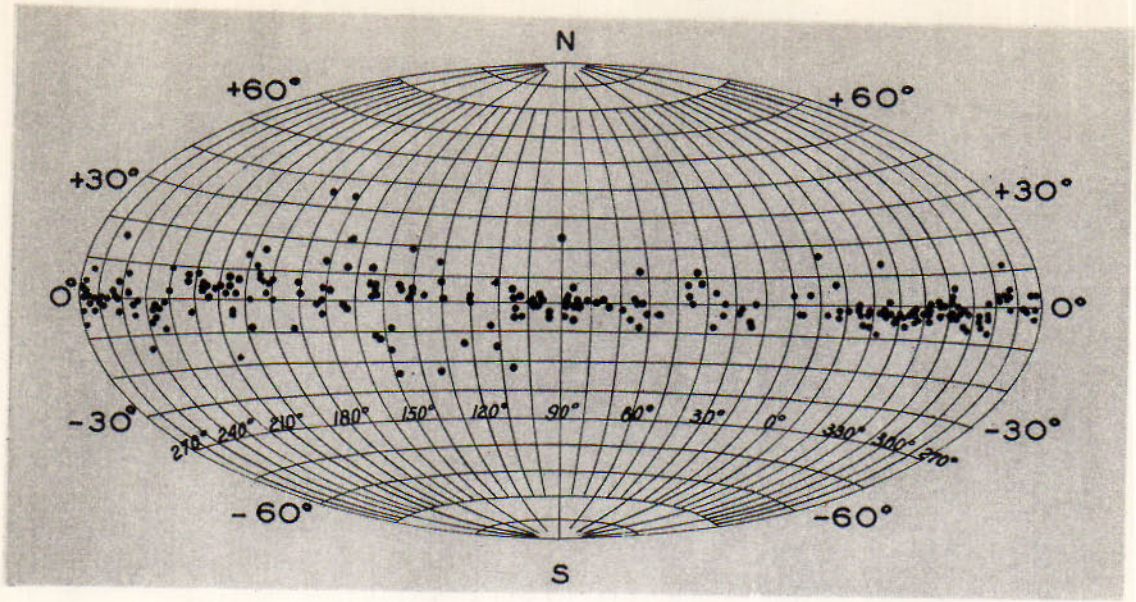


Fig. 668 — GALACTIC DISTRIBUTION OF OPEN CLUSTERS (SHAPLEY, 1930)

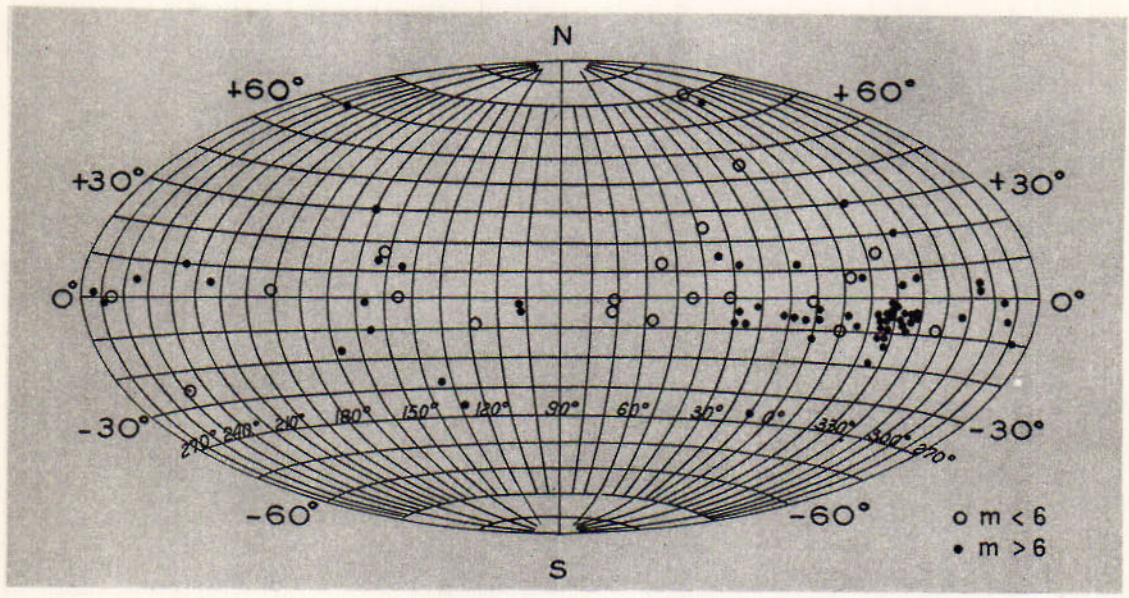


Fig. 669 — GALACTIC DISTRIBUTION OF NOVAE (C. P. AND S. GAPOSHKIN, 1936)

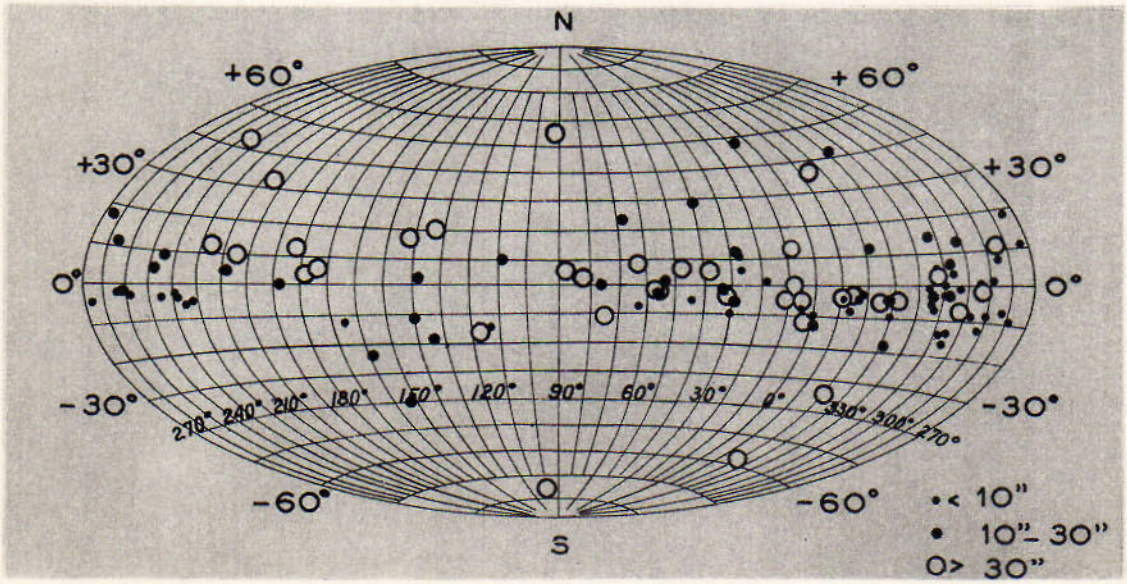


Fig. 670 — GALACTIC DISTRIBUTION OF PLANETARY NEBULAE (LUNDMARK, 1930)

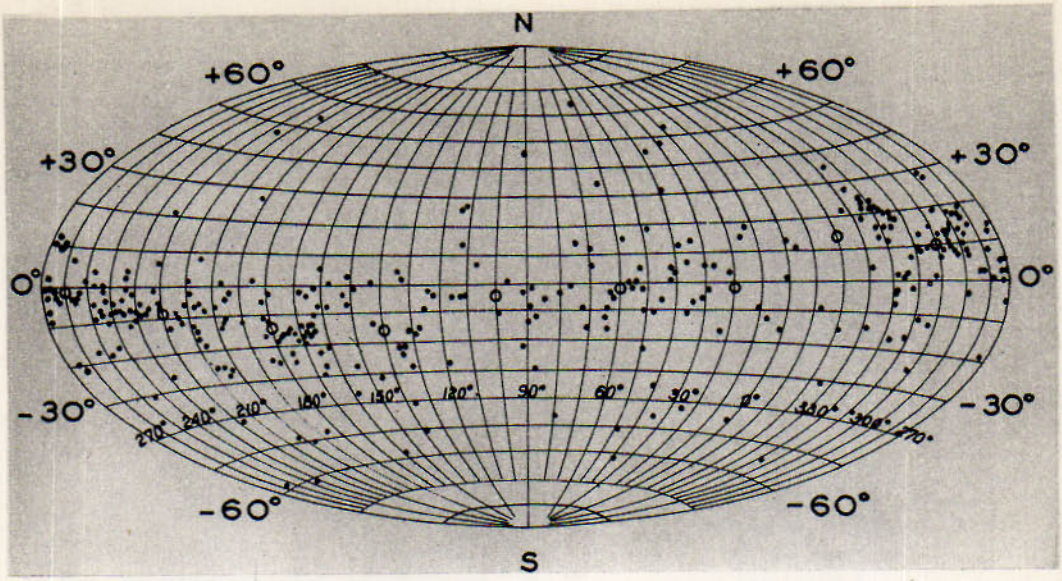


Fig. 671 — GALACTIC DISTRIBUTION OF TYPE B STARS BRIGHTER THAN MAGNITUDE 5, WHICH OUTLINES 'GOULD'S BELT' (LOCAL CLUSTER)

Fig. 672 — THE LOCAL STRUCTURE OF OUR GALAXY, AS DERIVED FROM RADIO-ASTRONOMICAL OBSERVATION OF THE 21-cm. HYDROGEN LINE. *(Photo, G. Westerhout, Sterrewacht te Leiden)*

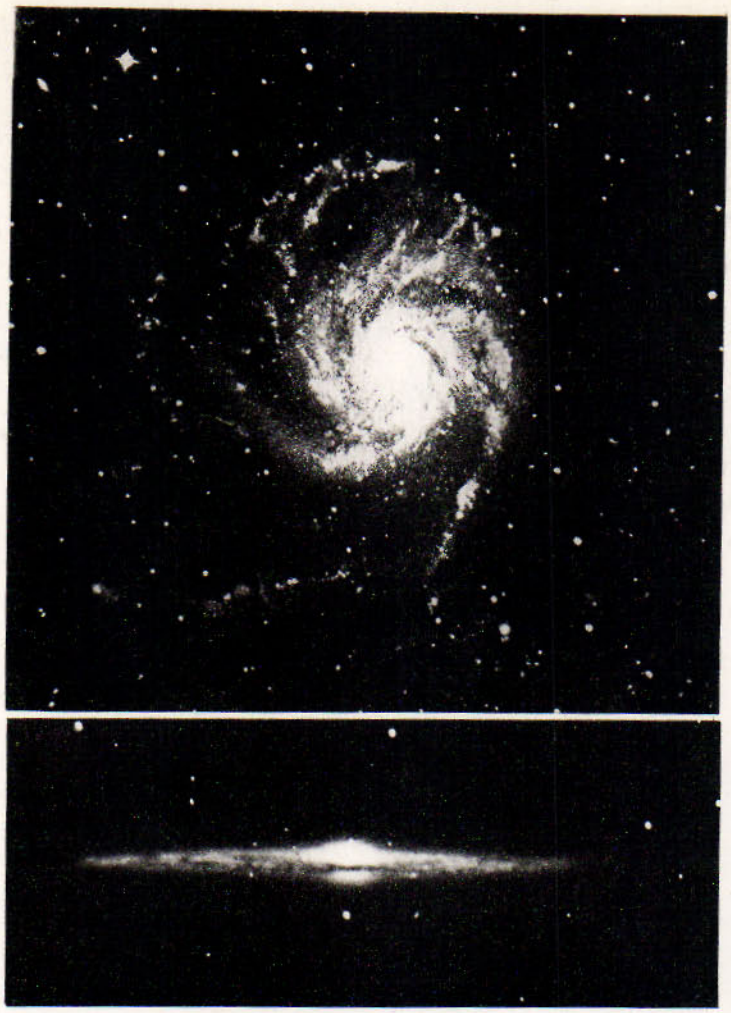


Fig. 673 — THESE TWO SPIRAL NEBULAE (TYPE *Sc*) — ONE SEEN IN PLAN AND THE OTHER EDGE-ON — GIVE SOME IDEA OF THE PROBABLE GENERAL APPEARANCE OF OUR GALAXY AS SEEN BY AN OBSERVER IN EXTRAGALACTIC SPACE. Above: M 101 in Ursa Major, Yerkes 24-inch telescope; below: NGC 4565 in Coma Berenices, Lick 36-inch telescope.

the centre outward, being no more than 1,000 to 2,000 lightyears in the vicinity of the Sun, at a distance of some 30,000 lightyears from the centre.

Another method of investigating the structure of the Galaxy, of more recent date, is based on the employment of the radio-waves or 'cosmic static' produced by the interstellar gas. Over the past 20 years communications engineers have frequently noticed when using the range of wavelengths of the order of metres that there was a certain amount of interference (or 'noise') obviously penetrating the ionosphere from outer space. The intensity of this noise was found to be maximum when the receiving antenna was directed towards the galactic plane. Moreover, if lines connecting points from which equal amounts of noise issued were plotted on a star-map, these 'isophotes' were found to agree remarkably well with the outline of the visible Milky Way, and the noise became most intense in the direction of Sagittarius.

The exact origin of this noise is still uncertain. A part of it may originate in stellar atmospheres, but the bulk is undoubtedly due to the motion of free electrons in the neighbourhood of the proton gas in interstellar space (i.e. the 'free-free' emission). It does, however, appear certain that — unlike light at optical frequencies — this radio noise can easily penetrate through the entire Galaxy, and should thus enable us, in the future, to ascertain the structure — i.e. the distribution of mass — of the central regions of our Galaxy, which remain invisible to us in our station in space because of the heavy interstellar absorption of ordinary light.

Much useful information has already been garnered by this method; some of the most recent data are shown in Fig. 672, and it is extensively dealt with in Chapter 19.

The results that have already been obtained allow us to sketch in the main outlines of the star system. But of equal interest is its detailed structure, particularly in the Sun's neighbourhood: do the stars thin out progressively towards the surface of the system, or are they grouped into secondary condensations, as the appearance of the Milky Way star clouds suggests? And if they are, how are these clouds distributed in space? So far, the difficulties imposed by the irregular interstellar absorption have prevented any definitive answers being given to these questions, although certain local features of galactic structure seem to be well established. The most important of these is the existence of a region of relatively high star density surrounding the Sun. This had already been suspected in about 1835 by J. Herschel, during his sojourn

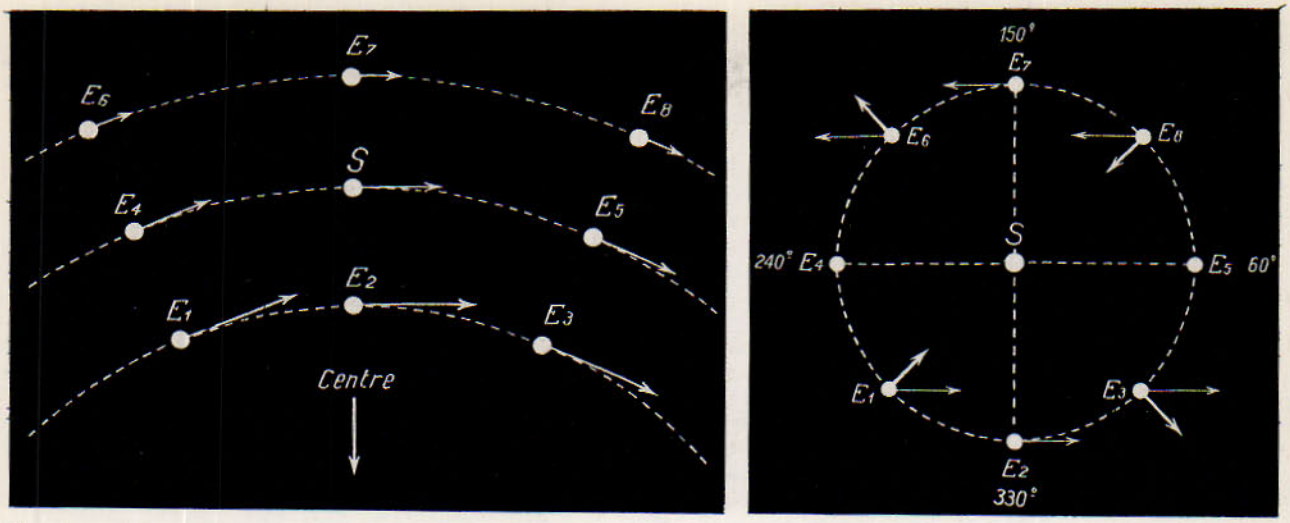


Fig. 674 — DIFFERENTIAL ROTATION OF THE GALAXY. Left: absolute spatial velocities; right: relative velocities, and radial velocities referred to the Sun.

at the Cape, and it was further investigated around 1880 by Gould, then Director of the Cordoba Observatory in Argentina. He drew attention to the fact that the brightest stars appear to be distributed round the star sphere in a zone which is sensibly inclined to the plane of the Milky Way, and suggested the existence of a flattened cluster in the Sun's vicinity, whose median plane is inclined at about 15° to that of the Galaxy. The centre of this so-called Local Cluster lies in the direction of Carina in about galactic longitude 270°.

The existence of the Local Cluster also seems to be indicated by the fact that the star density of space apparently decreases in every direction from the Sun — not only towards the poles of the Galaxy but even within its plane. The Dutch astronomer van Rhijn, for example, has shown that the density of B type stars is 6 or 7 times less at a distance of 1,000 lightyears than it is in the Sun's immediate vicinity. The objective existence of a well-defined Local Cluster is nevertheless strongly contested in some quarters, for although the star density falls off throughout the half of the Milky Way containing the anticentre its decrease is less evident in the directions of Carina and Cygnus, while in the direction of the galactic centre the interstellar absorption and the dark nebulae complicate matters to such an extent that no certain conclusions can be drawn.

Moreover, astronomers today — without giving up hope of disentangling, little by little, the local structure of the Galaxy — are trying rather to deduce it from comparisons with external galaxies, whose general layout and detailed structure can be observed without difficulty. Buried as we are in the interior of our own Galaxy, our situation is rather like that of someone lost in an insoluble maze.

And, indeed, the picture of the Galaxy which we have arrived at — with its central condensation, its thin section, and its dark clouds in the median plane — closely resembles the appearance of the extragalactic nebulae when viewed edge-on. Now the majority of these nebulae that are seen in plan (i.e. at 90° from the edge-on position) exhibit an altogether characteristic spiral structure. And the close similarity between the stellar population of the Galaxy and those of the extragalactic nebulae (globular clusters, O and B type stars and the diffuse nebulosity associated with them) suggests that the Milky Way should be classed as one of them — probably among the *Sb* or *Sc* spirals (see next chapter), which have a relatively small nucleus round which unwind enormous spiral arms, irregular in form and of complex structure. Continuing the analogy, it is generally supposed that the Sun is located in an outer arm of the galactic spiral, and that the Local Cluster is no more than a minor and accidental condensation in this arm, of a type often to be seen in the spiral nebulae.

This hypothesis rests on a number of different foundations. In particular, the presence of widespread galactic clouds in Carina and Cygnus, almost exactly opposite one another, and the distribution of the B type stars in the galactic plane (corrected as far as possible for the effects of interstellar absorption) hint at the existence of a spiral arm issuing from Carina and stretching past the Sun in the direction of Cygnus — i.e. lying in a direction which, as would be expected, is roughly at right angles to that of the centre of the Galaxy. It is also supported by star counts in the galactic clouds nearest to the nucleus, which in general show a density maximum separated from that in the Sun's vicinity by a wide zone of low density; for example, the great Scutum star cloud lies, according to recent estimates, at a distance of between 1 and 3 kiloparsecs, and is clearly separated from the Local Cluster by a region of very low star density situated at a distance of about 0.5 kiloparsecs from the Sun, the Local Cluster itself having in this direction a mean diameter of the order of 1 kiloparsec in the galactic plane. On the other hand, in the directions of Cygnus and Carina the star density seems to remain tolerably constant on either side of the Sun to distances of from 1 to 2 kiloparsecs; it even increases to a measurable extent here and there in the direction of Carina.

The most impressive evidence yet garnered, however, to support this conception of the Galaxy's structure, is of extremely recent date (1958). It rests upon the radio-telescopic observation of the 21-cm. line of neutral hydrogen, studied by the method fully described in Chapter 19. The data obtained by this technique have been combined to produce Fig. 672, which gives an impression of the Galactic System, as seen from the direction of the north galactic pole.

Although this photograph only gives the distribution of neutral hydrogen, since hydrogen is far and away the commonest element in the Galaxy it is safe to assume that the bright stars and hot gas clouds follow the same pattern. These stars and gas clouds are those that we see in the spiral formation made familiar to us by the study of other

galaxies; in total, however, they contribute only 10% or so of the mass of the entire Galactic Systems. The balance is made up of fainter stars which do not make an appreciable contribution to the light emitted.

From observations of the 21-cm. spectral line both the density and distance of these hydrogen clouds can be determined. The distance from the Sun (denoted in the picture by ⊙) to the centre of the system (+) is 27,000 lightyears or 8,200 parsecs. The system rotates in a clockwise direction. Moving with a velocity of 220 k.p.s. approximately, the Sun takes something like 230 million years to complete one revolution.

These observations, which constitute the most remarkable so far of the actual structure of our Galaxy, were made by the Netherlands Foundation for Radio-astronomy in Leiden, and the Division of Radiophysics, Commonwealth Scientific and Industrial Research Organisation, Sydney, Australia; to these two groups of astronomers we owe our modern knowledge of the spiral structure of the Galaxy as a whole.

Finally, the comparison of the apparent brightness of the Galaxy in the Sun's neighbourhood, as it would appear to an observer in extragalactic space (derived from star counts) with that of the spiral nebulae, emphasises the very excentric position of the Sun within the Galaxy. We shall have more to say on this point in the next chapter, when the great neighbouring galaxies are compared with our own.

There is still, however, one extremely important feature of the Galaxy which we have not yet mentioned, save in passing. This — its rotation as a single unit — has only been established during the past 30 years. By this means, it has been possible to calculate the total mass of the Galaxy.

Rotation and mass of the Galaxy. — The high degree of flattening of the Galaxy, and its resemblance to the spiral nebulae, both suggested very forcibly that it is in rapid rotation about an axis passing through its centre. If this is so, every star must describe a circular or elliptical orbit about this centre, in or parallel to the median plane of the Galaxy.

But if this were so, one would not expect the apparent motions of the stars to be random; on the contrary, they should follow certain preferential directions, dictated by the rotation of the whole system. The existence of such preferred directions among the motions of the stars had, in fact, been demonstrated as long ago as 1904 by the Dutch astronomer Kapteyn, who had shown that stellar proper motions, when corrected for that of the Sun, form two main streams through space. These are flowing towards opposite regions of the sky (Orion and Scutum), although both lie in the plane of the Milky Way. Some years later, the German astronomer K. Schwarzschild presented the same conclusion in a different form: he found that the preferential directions of the motions of the stars were respectively towards Sagittarius (longitude 340°) and Orion — approximately the centre and the anticentre. But at that time the reason for the preference was not known.

Another curious fact, concerning stellar radial velocities was discovered by the American astronomer L. Boss in 1918, and was studied by (in particular) the Dutch astronomer Oort and the American Stromberg between 1925 and 1930. Stars with radial velocities lower than about 40 m.p.s. (after correction for the Sun's motion) were found to be moving more or less at random in every direction within the galactic plane, as was expected. But the distribution of stellar velocities greater than 40 m.p.s. is extremely asymmetrical: the motions of nearly all high-velocity stars are directed towards the region between galactic longitudes 140° and 340° — i.e. the southern stretches of the Milky Way and Taurus to Sagittarius — and practically none in the opposite direction towards the Scutum—Cygnus—Auriga region. This asymmetry of stellar motions is thus distributed on either side of a mean direction whose galactic longitude is 230°, situated at right angles to the direction of the galactic centre.

The explanation of these facts — known respectively as Kapteyn's star streaming and Strömberg's asymmetrical

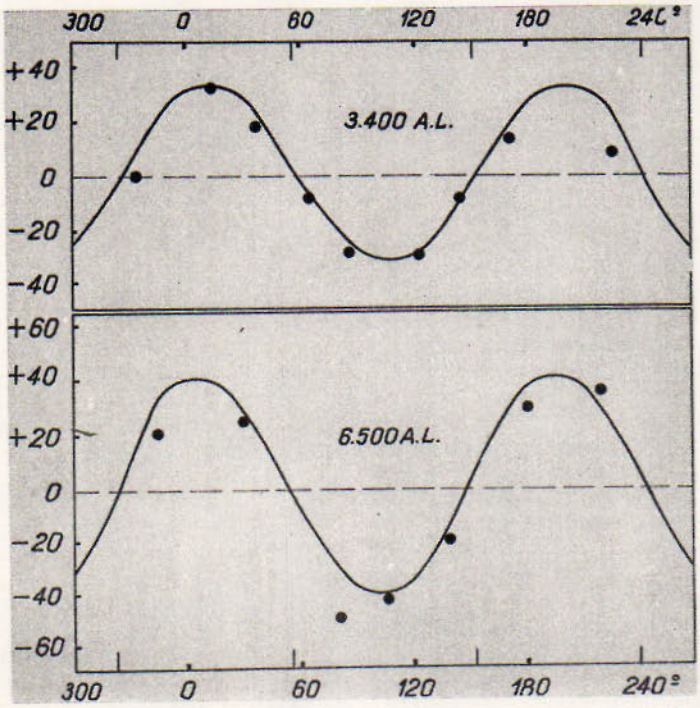


Fig. 675 — DIFFERENTIAL ROTATION OF THE GALAXY. Radial velocities of two groups of Cepheids (after Joy).

star streams — was provided in 1927 by the Swedish astronomer Lindblad, who showed that the Galaxy as a whole is rotating, and by Oort, who detected its differential rotation (of which we shall speak later).

Strömberg had in fact shown that the degree of asymmetry of the radial velocities depended on the sort of object considered: in particular, it is very marked in the case of the globular clusters and extragalactic nebulae. Now we know that although the system of the globular clusters is closely connected with the Galaxy, it is nevertheless very little flattened: in fact the volume of space occupied by the globular clusters is almost spherical. This indicates that the rotation of the globular cluster system must be virtually negligible compared with that of the majority of the galactic stars (including the Sun), which belong to a system that is both very concentrated and very flattened. Consequently the globular clusters provide a frame of reference, more or less fixed in space, to which the rotation of the Galaxy can be referred. Now the globular clusters as a whole appear to be moving towards the constellation Argo (longitude 240°) at a mean velocity of about 170 m.p.s.; and just as the apparent general drift of the nearer stars towards the Antapex is only a reflection of the Sun's own motion towards the Apex (p. 308), so this general apparent movement of the globular clusters must result from the Sun's real motion in the opposite direction — i.e. towards longitude 60° in Cygnus. The measured radial velocities of the nearer extragalactic nebulae lead to the same conclusion.

Thus the Sun partakes of a rapid rotational motion about the distant galactic centre — which incidentally provides still further evidence for the latter's location in the direction of Sagittarius. Since the direction of this motion is almost exactly at right angles to that of the galactic centre, the Sun's orbit must be very nearly circular. We can see, also, that this motion will override that of the Sun towards the traditional Apex (recognised from a much earlier date) at a velocity of 12 m.p.s. relative to the nearer stars; these stars, as a whole, are thus moving in practically the same direction at a velocity of 170 − 12 = 158 m.p.s., for the classical Apex, near Vega (α Lyrae) is at no great distance from the galactic apex, near Deneb (α Cygni).

Granted this, Kapteyn's star streaming and Strömberg's asymmetry of stars with large radial velocities are immediately explicable. A considerable spread of stellar velocities is in fact observed among the Sun's neighbours, amounting to several tens of m.p.s. relative to the Sun. This results from the fact that not all the stellar orbits are perfectly circular, some being to a greater or smaller degree elliptical. For at any given distance from a centre of gravitational attraction, the velocity of a body moving in an elliptical orbit is greater than that of one moving in a circular orbit: we have already seen, for example, that the Earth's velocity in its almost circular orbit is 18 m.p.s., whereas comets and meteors at the same distance from the Sun, but moving in very elliptical orbits, have significantly higher velocities of the order of 25 m.p.s. This being so in the case of the stars also, some will be moving slightly towards the outside of the system, relative to the Sun, and others towards the centre. This will appear to us under the guise of a tendency to stream in one or other of two opposite directions: in other words, Kapteyn's star streams.

In the same way, the asymmetry of large stellar radial velocities is easily explained. We know that the stars with large radial velocities (short-period Cepheids, for example) are for the most part not strongly condensed towards the galactic plane; they thus constitute comparatively unflattened systems whose common rotational motion is not very rapid. It must therefore be concluded that these apparently high-velocity stars have in fact slower motions about the galactic centre than the Sun, and they only appear to be moving so fast because of the Sun's high velocity relative to them. But then, since the Sun is moving more rapidly than these stars towards the galactic apex in longitude 60°, their motion relative to the Sun will appear to lie in the opposite direction — i.e. the galactic hemisphere centred on the Antapex. Hence Strömberg's asymmetry.

It may nevertheless be asked why the rotational velocities of no stars exceed that of the Sun by more than about 35 m.p.s. — why, in fact, none of the stars in our vicinity have velocities about the galactic centre of, say, 250 m.p.s. The answer is that the gravitational field of the Galaxy is not strong enough to retain any star moving as fast as this: it would exceed the so-called parabolic velocity, or escape velocity, which in the Sun's vicinity is 200 m.p.s. If there were once any stars with such velocities they have long ago escaped from the Galaxy and are now lost in the immensity of extragalactic space. In the same way, comets and meteors which pass the Earth at velocities exceeding 26 m.p.s. are moving in hyperbolic orbits and are not permanent members of the Solar System.

Up to this point we have been arguing as though the Sun's stellar neighbours were being displaced *en bloc* by the rotation of the Galaxy. In reality, however, their velocities will vary with their distance from the galactic centre. We have, indeed, already seen how the planets revolve about the Sun (in which virtually the whole mass of the Solar System is concentrated) with velocities whose linear and angular values decrease with increasing distance: the orbital velocity of the Earth, for example, is 18 m.p.s., or roughly 1° of its orbit per day, while that of Venus, situated nearer to the Sun, is 22 m.p.s., and in the same interval carries it through an arc of over 1.5°; Mars, on the other hand, covers only 0.5°, at a linear velocity of 15 m.p.s.

The stars would be expected to show an analogous effect, though it will be less marked in their case since the whole mass of the Galaxy is not condensed in a single central body but is spread throughout the whole extent of the system (though it is admittedly concentrated towards the centre). This effect of differential galactic rotation was first detected by Oort, in 1927, and has been the subject of intensive study since that time, notably by J. S. Plaskett in Canada and H. Mineur in France.

It is most easily demonstrated by the radial velocities of the stars (see Fig. 674a). To simplify the problem, let us suppose that all the stars are moving in concentric circular orbits. Those stars, such as E_1, E_2 and E_3, which are nearer than the Sun to the galactic centre, will be moving more rapidly than the Sun; the velocities of E_4 and E_5, at the same distance as the Sun from the galactic centre, will be equal to the Sun's; and stars E_6, E_7 and E_8 will be moving more slowly than the Sun. Remembering the definition of radial velocity (p. 307), it will be clear that the line-of-sight velocities of stars E_2 and E_7, relative to the Sun, will be zero, since their motions are perpendicular to the line of sight; so will those of E_4 and E_8, whose orbital velocities are the same as the Sun's. Stars E_1 and E_8, on the other hand, will appear to be approaching the Sun with velocities that are equal to the projection of the difference of their velocities on the line of sight, while stars E_3 and E_6 will seem to be receding from the Sun at the same velocities.

In reality each of the stars E_1, E_2 . . . . . E_8 must be replaced by a group of at least several score members, in each of which the individual motions of its members will tend to cancel one another out, revealing the motion of the group as a whole, due to galactic rotation. It is found

that under these circumstances the radial velocities of the Sun's neighbours are distributed in the form of a sine curve (Fig. 674b), being nil in four mutually perpendicular directions, one of which must be that of the galactic centre. Furthermore, since differences of rotational velocity should increase in proportion with the separation between the stars' orbits and that of the Sun, the effect of differential rotation would be expected to increase in step with the distances of the star groups considered — so long, at least, as these distances remain small compared with the radius of the Galaxy, i.e. up to 1 or 2 kiloparsecs, which in any case corresponds roughly with the limit to which we can at present reach.

Fig. 675, relating to two groups of Cepheids studied by Joy at Mount Wilson, shows that (within the limits of the accuracy of the observations, and allowing for individual discrepancies) the theoretical predicions of Oort are magnificently confirmed by observation; and in particular, that the radial velocities are indeed zero in the region of longitude 325° — that of the galactic centre. The coefficient of proportionality between the amplitude of the Oort effect and the distance of the stars considered (known as Oort's constant) has been found from numerous investigations of different types of star to amount to 10 m.p.s. per kiloparsec. In other words, the rotational velocity of stars about the galactic centre increases or decreases by 10 m.p.s. relative to the Sun, for every kiloparsec (3,200 lightyears) from the Sun towards the interior or towards the exterior of the Galaxy. It will be realised that it is not practicable to detect this effect among the nearby stars which were employed to determine the classical Apex of the Sun's path.

One very interesting consequence of the differential rotation is that it provides a new method for determining the distances of certain remote groups of stars, and hence their absolute magnitudes. It is only necessary to determine the amplitude of their Oort effect and to divide this by 10 to have their distance in kiloparsecs, the derived value being independent of interstellar absorption.

It should be mentioned that evidence for differential galactic rotation has also been provided by measures of the radial velocities of the interstellar gas whose characteristic lines appear in the spectra of distant high-temperature stars, as already described. These lines give evidence of an Oort effect, increasing with distance, but its amplitude is only half that of the stars. This proves conclusively that the absorption occurs all along the light-path from the stars to the Earth — so that its mean value is realised at the midway point — and not only in the star's immediate vicinity, as was once thought. The interstellar cloud thus partakes of the general galactic rotation as shown by the stars — which is in no way surprising, in view of its strong concentration about the galactic plane.

Oort's effect can also, though with greater difficulty, be demonstrated by stellar proper motions, and the distance of the centre of rotation can be derived by comparing this with the Oort effect determined from the study of radial velocities. The figure so obtained is 9 to 10 kiloparsecs, which agrees very well with that indicated by the study of globular clusters.

In the same way it can be calculated that the period of rotation at the Sun's distance from the centre is about 2×10^8 years; this period, required by the Sun and the stars in its vicinity for one complete revolution about the galactic centre, is sometimes referred to as the 'cosmic year'. The annual angular rotation of the Galaxy in the Sun's neighbourhood thus amounts to only 0.006″.

The theory of galactic rotation developed by Lindblad furthermore enables us to estimate the mass of that part of the Galaxy lying within the Sun's orbit: the figure so obtained is 1.6×10^{11} times the mass of the Sun, about two-thirds of this being concentrated in the region of the nucleus, and one-third in or between the spiral arms. The mass of the peripheral region outside the Sun's orbit must be less than this, so that the total mass of the Galaxy may be put at about 2×10^{11} Suns. This mass is probably more or less equally distributed between the stars and the interstellar medium; however, this estimate of the relative masses of the stars and the diffuse galactic material is probably subject to revision, since it is not given directly by observation, and, moreover, recent work on the rotation of spiral nebulae has led to somewhat different results.

We must now turn our attention to these distant galaxies, whose realm extends from the borders of our own stellar system to the furthermost frontiers of the Universe so far explored.

Fig. 676 — THE GREAT NEBULA IN ANDROMEDA. *(Photo, Barnard, Yerkes Observatory, 10-inch telescope)*

CHAPTER 17

THE EXTRAGALACTIC NEBULAE AND THE EXPANSION OF THE UNIVERSE

The exploration of our Galaxy has revealed a multitudinous assemblage of bodies of every description and often of impressive appearance, which together constitute a vast whirlpool in which the Sun is no more than an imperceptible point, lost among the thousands of millions of its kind. Coming at last to the frontiers of this system, the insatiable human spirit of enquiry would be a little disappointed if it were to be confronted with nothing beyond but empty space — dark, glacial and without end.

Extragalactic nebulae, and the 'island universe' theory. — However, as soon as the status of our Galaxy as a physical unit with finite boundaries was suspected, it appeared logical to infer the existence of similar systems scattered throughout space far beyond its confines. Kant, indeed, had already suggested, two centuries ago, that these immensely distant sister-galaxies are in fact the objects termed 'nebulae' (appearing as pale, elliptical or hazy patches of light), whose nature was energetically discussed by astronomers. Kant's theory was confirmed at the end of the eighteenth century by William Herschel who, with the aid of his giant reflectors, discovered thousands of new nebulae; from that time, the existence of other stars systems, scattered everywhere throughout the Cosmos, was generally accepted. The name 'island universe', which was given to them by the German, Humboldt, in 1850, was immediately taken up and has been employed ever since.

Even at that time it was widely held that it only needed sufficiently powerful telescopes for many, if not all, of these nebulae to be resolved into stars — just as Herschel had succeeded in resolving the globular clusters, with which they appeared to be comparable. And, no doubt, inspired by this hope, Lord Rosse believed that he had succeeded in resolving many nebulae into stellar points with his great 72-inch reflector.

But an entirely new light was thrown on the question when, in 1864, Huggins showed by means of spectroscopic analysis that many of the planetary and diffuse nebulae (such as that in Orion) are in fact composed, not of stars, but of highly rarefied gas whose spectrum consists of isolated bright lines; one of these, a particularly bright line in the green, is responsible for the characteristic colour of these nebulae. Others, however (like the

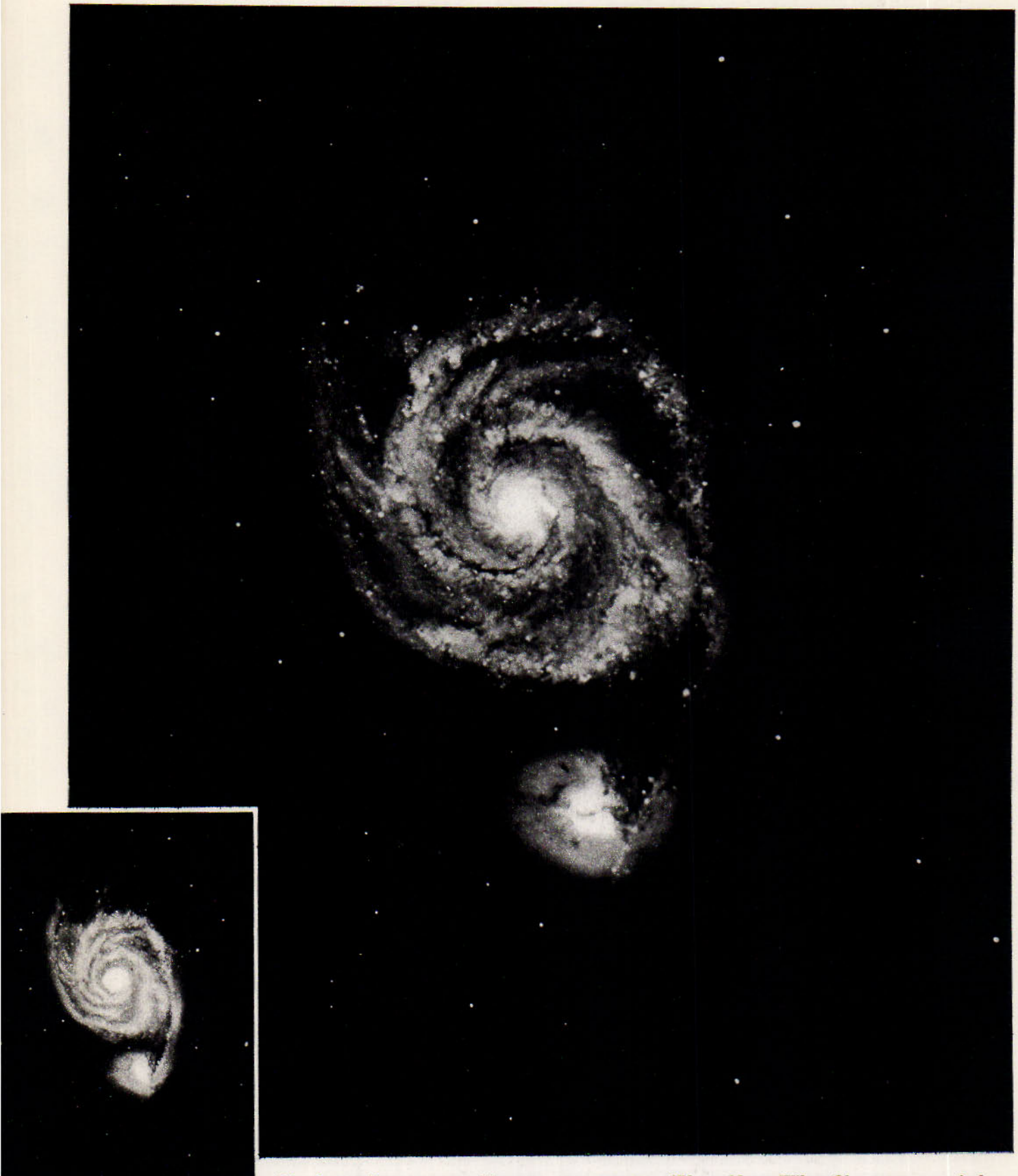


Fig. 677 — SPIRAL NEBULA M 51 IN CANES VENATICI. *(Photo, Mount Wilson Observatory, 100-inch reflector, exposure 3h; 15 May 1926)*
Inset: THE SAME NEBULA PHOTOGRAPHED WITH A 20-INCH TELESCOPE. *(Photo, I. Roberts, exposure 2h 25m; 25 May 1896)*

Fig. 678 — THE GREAT SPIRAL NEBULA IN ANDROMEDA (M 31) AND ITS TWO ELLIPTICAL COMPANIONS (M 32, BELOW THE CENTRE OF M 31, AND NGC 205, UPPER RIGHT). *(Photo, G. W. Ritchey, Yerkes Observatory, 22-inch telescope; 1901)*

Fig. 679 — CENTRAL REGION OF THE ANDROMEDA NEBULA, SHOWING THE NUMEROUS DARK CLOUDS OF OBSCURING MATTER ANALOGOUS TO THE DARK NEBULAE OF THE GALAXY. *(Photo, Mount Wilson Observatory, 100-inch reflector, exposure 9h; 1920)*

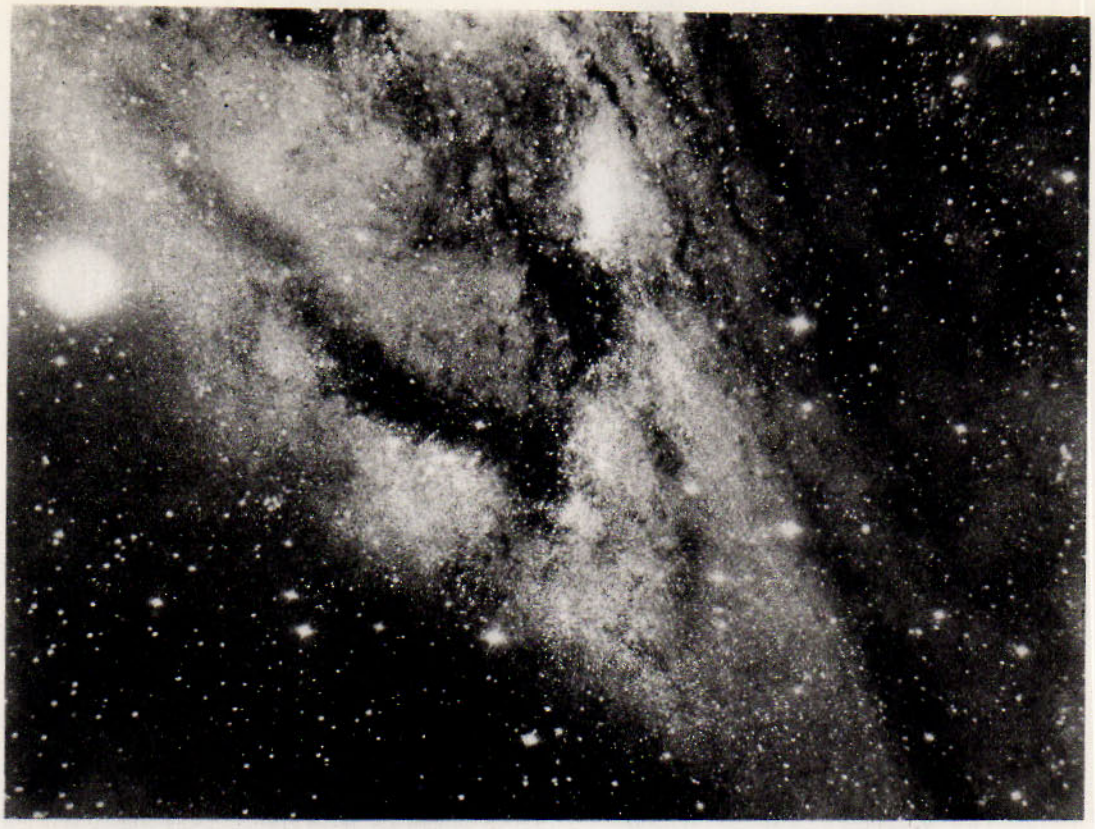


Fig. 680 — THE SOUTH-WEST EXTREMITY OF THE GREAT ANDROMEDA NEBULA RESOLVED INTO INDIVIDUAL STARS. *(Photo, Mount Wilson Observatory, 100-inch telescope, exposure 2h; 24 August 1925)*

great nebula in Andromeda), are almost pure white in colour, and their spectra are more or less continuous — i.e. of the same general type as that of the Sun. That fact alone, however, was not enough to establish their stellar character, for the existence of dust clouds, shining by reflected light and giving continuous spectra (e.g. the Pleiades), was demonstrated by Slipher in 1912; and, furthermore, measurements of the parallax of the

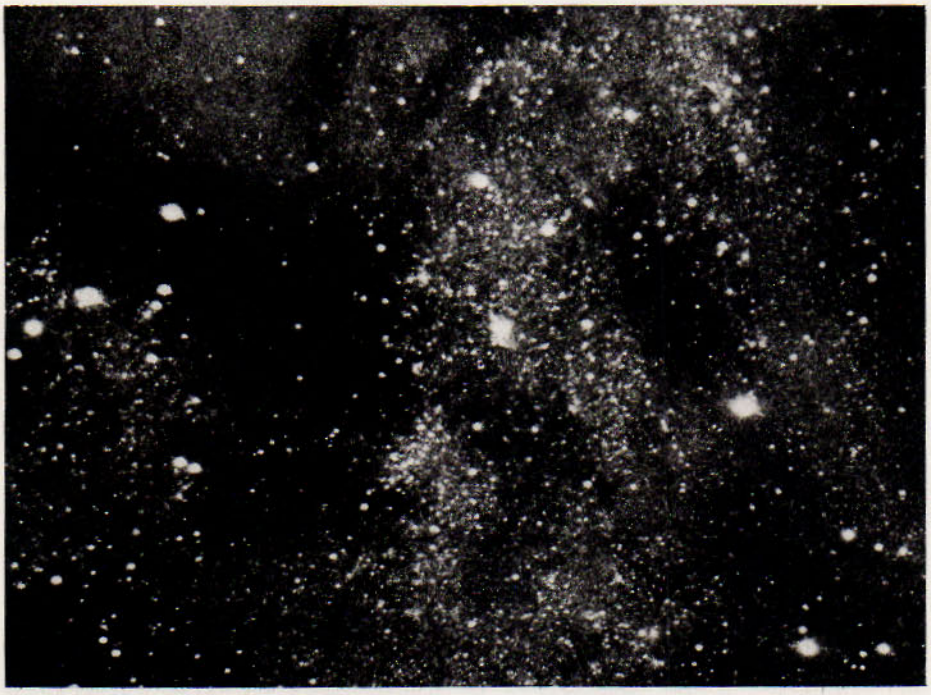


Fig. 681 — RESOLUTION INTO SEPARATE STARS OF A SMALL REGION OF THE NEBULA M 31 IN ANDROMEDA. (Enlargement of part of the preceding photograph.)

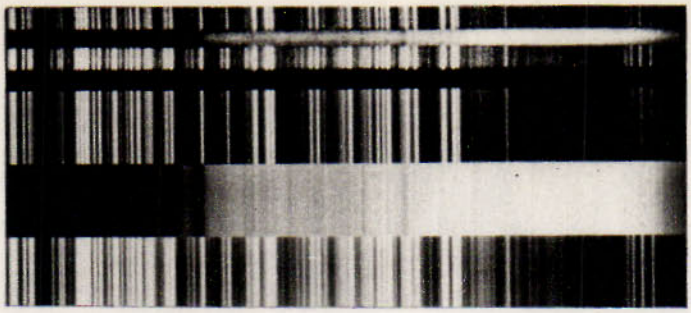


Fig. 682 — THE SPECTRUM OF THE CENTRAL REGIONS OF THE ANDROMEDA NEBULA CLOSELY RESEMBLES THAT OF THE SUN. *Centre*, the G band; *left*, the H and K lines of ionised calcium; *above*, the original spectrum; below, the same spectrum enlarged. *(Lick Observatory)*

Andromeda nebula that were made at about the same time seemed to place it well within the frontiers of the Milky Way.

The question of the status of the 'white nebulae' was finally and definitively answered by Hubble, in the mid-1920s. As a result of Ritchey's accidental discovery of first one and then several novae in the Andromeda nebula (made at Mount Wilson in 1917), Hubble had undertaken a systematic study of this and other nebulae with the great Mount Wilson telescopes. Photographs taken with the 100-inch reflector finally effected the partial resolution of the Andromeda nebula into myriads of very faint stars; among these were recognised numerous variables, including Cepheids. The latter, by the period-luminosity relation, immediately allowed the distance of the nebula to be deduced: it was found to be in the neighbourhood of one *million* lightyears. Thus, at last, the stellar composition of these nebulae, their general similarity to the Galaxy, and their location far beyond the bounds of our own stellar system, were incontestably established. From that time our knowledge of the 'island universes' has progressed by leaps and bounds — mainly as a result of work done at Mount Wilson. Today, little more than 30 years later, the known Universe extends to unimaginable distances beyond the frontiers of the Galaxy, the limit to which existing instruments can penetrate space being more than 2×10^9 lightyears. This, then, is the vista of space which we can now describe, in the light of these epoch-making investigations.

But before doing so, it might be as well to say something about the nomenclature of the extragalactic nebulae. The early telescopic observers — seeing these objects as faint, small and ill-defined patches of light — very naturally called them 'nebulae', the name by which they characterised all objects having this particular appearance; the usage has been hallowed by tradition and is current to this day, even though it is now realised that they have nothing in common with the diffuse nebulae of our own Galaxy. Hubble suggested that this designation should be made more explicit by the qualifying epithet 'galactic' or 'extragalactic', as the case might be, and the term 'extragalactic nebula' is in fact the one most commonly used today. On the other hand Shapley, at Harvard, suggested the term 'external galaxies', or simply 'galaxies', since they are recognised as being analogous to the Galaxy; but this name — which could on occasion lead to confusion with the Galaxy — has been less widely adopted than Hubble's.

Catalogues and counts of the nebulae. — Only three extragalactic objects are visible to the naked eye: the two patches of nebulosity in the southern hemisphere,

THE BRIGHTEST EXTRAGALACTIC NEBULAE ($m<10$)
(after H. Shapley and A. Ames, and M. Humason)

NUMBER	CONSTELLATION	TYPE	MAG. (PHOTOG.)	APPARENT DIMENSIONS (SECONDS OF ARC)	RADIAL VELOCITY (m.p.s.)
NGC 55	Sculptor	S	7.8	25 × 3	
NGC 224 (M 31)...	Andromeda	*Sb*	5 ±	160 × 40	− 200
NGC 221 (M 32)...	Andromeda	E2	9.5	2.6 × 2.1	− 200
NGC 253	Sculptor	*Sc*	7.0	22 × 6	
Nubecula Min.	Toucan	I	1.5	216 × 216	± 100
NGC 598 (M 33)...	Triangulum	*Sc*	7.8	60 × 40	− 110
Nubecula Maj.	Norma	I	0.5	432 × 432	± 170
NGC 3031 (M 81)..	Ursa Major	*Sb*	8.9	16 × 10	− 20
NGC 3034 (M 82)..	Ursa Major	I	9.4	7.0 × 1.5	+ 180
NGC 3115	Sextans	E7	9.8	4.0 × 1.0	+ 350
NGC 3627 (M 66)..	Leo	*Sb*	9.9	8.0 × 2.5	+ 400
NGC 4594	Corvus	*Sa*	8.1	7.0 × 1.5	+ 710
NGC 4631	Can. Venat.	*Sc*	9.6	12.0 × 1.2	
NGC 4736 (M 64)..	Can. Venat.	*Sb*	9.0	5.0 × 3.5	+ 180
NGC 4826	Coma	*Sb*	8.0	8.0 × 4.0	+ 90
NGC 4945	Centaurus	S	9.2	11.5 × 2.0	
NGC 5128	Centaurus	I	7.2	10.0 × 8.0	
NGC 5236 (M 83)..	Hydra	*Sc*	8.0	10.0 × 8.0	+ 300
NGC 5457 (M 101).	Ursa Major	*Sc*	9.0	22 × 22	+ 200
NGC 7793	Sculptor	S	9.7	6.0 × 4.0	

resembling detached pieces of the Milky Way, known as the Magellanic Clouds* after the famous navigator who observed them during the course of his circumnavigation; and, in the northern skies, the great nebula in Andromeda, which lies close to the star ν Andromedae. It appears as a small, diffuse, elongated patch of light of the 4th magnitude, and was first described by Simon Marius in 1612.

Most of the other 'nebulae' discovered during the 17th and 18th centuries were galactic objects — diffuse nebulae, and globular and open clusters, which were indistinguishable from one another in the telescopes of that epoch; indeed, up to the beginning of the 20th century all the great catalogues of these objects listed nebulae and clusters quite impartially.

The first astronomer to compile a systematic list of nebulae and clusters was the Frenchman, Charles Messier, whose aim in doing so was merely to avoid confusing them with comets, of which he was an ardent and successful seeker (he discovered no fewer than 21 with a small telescope that he set up on the Tower of Cluny, in Paris). His catalogue, published in 1784, listed 103 objects, of which about 50 are extragalactic.

The numbers in Messier's catalogue are still widely used for designating the brighter nebulae and clusters: the Andromeda nebula, for example, is M 31, its smaller neighbour is M 32, and the very fine spiral (also visible with a small glass) in the nearby constellation of Triangulum is M 33. The table (p. 399), which lists the 20 brightest extragalactic nebulae, gives Messier's number for those objects that were included in his catalogue.

More comprehensive catalogues were compiled during the 19th century, notably by William Herschel and his son, John, from their own observations. In 1890, J. L. E. Dreyer's *New General Catalogue* (NGC) was published, in which the lists of the Herschels and of other observers were collated. The NGC was brought up to date to the year 1907 by the further publication of two *Index Catalogues* (IC) in 1895 and 1910. Altogether, more than 13,000 objects, including over 12,000 extragalactic nebulae, figure in these catalogues (7,840 in the NGC, and 5,386 in the IC), and their numbers are the universally employed nomenclature for the nebulae. Thus the Andromeda nebula, M 31, is also NGC 224, M 32 is NGC 221, M 33 in Triangulum is NGC 598, etc.

Since the introduction of photography the number of recorded extragalactic nebulae has increased so rapidly that to catalogue them individually would be as impracticable as useless. Nevertheless, since the earlier catalogues were by no means homogeneous or mutually comparable, the need of valid statistics has in recent years necessitated the making of complete photographic surveys of the extragalactic nebulae to given magnitude limits. Thus a census of nebulae brighter than magnitude 13 was taken at Harvard College Observatory and published in 1932, containing 1,249 objects whose distances are in general less than 10^7 lightyears; this was followed by a still more important compilation of over 500,000 nebulae to magnitude 18, lying within a sphere whose radius is some 10^8 lightyears.

Such catalogues are not only useful for enabling the astronomer to locate and identify such-and-such a nebula: their importance lies in the fact that they provide the raw material for statistical studies of the spatial distribution of the extragalactic nebulae, their tendency to congregate in clusters, etc. — i.e. of the structure of the whole extragalactic ensemble, sometimes known as the Metagalaxy. It may be said at once that these investigations have revealed no progressive thinning out or concentration of the extragalactic nebulae, such as might suggest that our probings of space are approaching the edge or the centre of a super-system analogous to the Galaxy. Numerous local and irregular clusterings of nebulae have indeed been found, but the large-scale distribution of the extragalactic nebulae appears to be sensibly uniform to the extreme limit that our instruments can reach.

Another use to which these surveys of the extragalactic nebulae have been put is the investigation of interstellar absorption within the Galaxy, as described in the last chapter.

The statistical uniformity of their distribution in depth is particularly clearly shown by the nebulae counts to specific magnitude limits that have been made at the Lick and Mount Wilson Observatories: after the necessary corrections have been made, these show that down to magnitude 21 (i.e. to distances of the order of 5×10^8 lightyears), their numbers increase regularly in proportion to the volume of space plumbed by successive counts — and exactly in the proportion of 3.98:1 per magnitude (see p. 364). These are thus no indications of a limit to the Metagalaxy within this distance, nor can there be any appreciable absorption of light in intergalactic space — unlike intragalactic space, where interstellar absorption combined with the thinning out of the stars towards the borders of the stellar system produces, as we have seen, a proportional increase of numbers per magnitude significantly less than 3.98, and, furthermore, decreasing steadily as the magnitude increases.

This phenomenon is particularly noticeable in the vicinity of the galactic poles: down to about magnitude 13, for example, stars are much more numerous than nebulae, which are widely scattered and appear only occasionally on the photographs; but as one proceeds to lower and lower magnitudes, the number of stars increases more and more slowly while that of the nebulae increases steadily (being nearly quadrupled at each successive magnitude), with the result that at magnitude 21, the limit of the Mount Wilson 100-inch telescope, the plates shows as many nebulae as stars.

All the same, the demonstration of this uniformity of spatial distribution involves the application of certain corrections to the crude observational data (the nebular counts), and these are both fairly considerable and not very well determined; hence the conclusions that can be drawn from the counts are somewhat uncertain. We shall return to this point when discussing the problems raised by the so-called expansion of the universe.

These corrections are of no great moment so long as the most distant nebulae (fainter than magnitude 18) are ignored. No corrections are required, for instance, in the case of the nebulae brighter than this magnitude which figure in the Harvard catalogues. Under these circumstances there is a very marked asymmetry between the two galactic hemispheres: if the great clusters of nebulae are disregarded, there are from 30% to 40% more nebulae in the northern hemisphere than in the southern, although this difference decreases with increasing magnitude and beyond magnitude 18 vanishes altogether. Large-scale variations of the same order (likewise discovered by Shapley) are evident when we consider the distribution perpendicular to the line of sight over comparable distances — i.e. about 2×10^8 lightyears — in either hemisphere. But the nature and significance of these large-scale discrepancies are still unknown. In any case, if a metagalactic system does exist, and bears the same relation to the extragalactic nebulae as the Galaxy does to the stars, then it certainly extends, in all directions, beyond the limits that we can explore today; and, in the absence of any evidence to the contrary, it is justifiable to equate it with the Universe as a whole.

* Or the Nubecula Major and Minor.

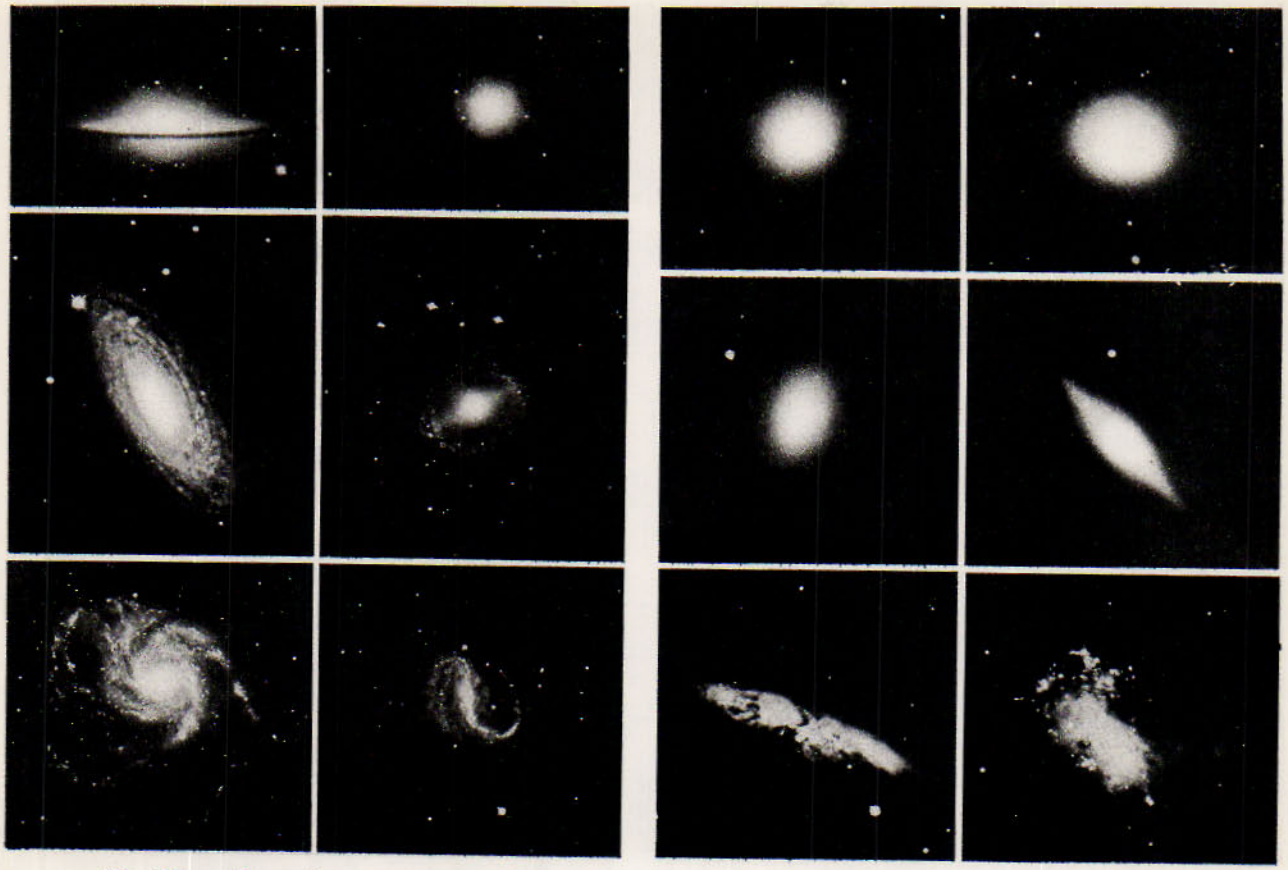


Fig. 683 — HUBBLE'S CLASSIFICATION OF THE EXTRAGALACTIC NEBULAE. *Left:* normal and barred spirals; *right:* elliptical and irregular nebulae.

The domain explorable with the Mount Wilson 100-inch reflector is something like 10^9 lightyears in diameter* (10^{22} miles, or 10,000 times the diameter of the Galaxy). Within this region of space about a million nebulae brighter than the 18th magnitude have been photographed, and it is estimated that the total population down to magnitude 21 (taking into account those that are hidden from us by the Milky Way) is of the order of 10^8 nebulae, distributed more or less evenly.

Appearance and classification of the extragalactic nebulae. — As seen in the small telescope of the early observers, the extragalactic nebulae appeared as faint, diffuse patches of light, circular or elliptical in shape, brighter at the centre than towards the edges, but showing no trace of internal structure. This only began to be observed with Herschel's powerful telescopes, and still more with that of Lord Rosse towards the middle of the 19th century. This instrument — a 72-inch reflector — showed him, to his amazement, that many of the nebulae are in fact magnificent whirlpool structures with spiral arms unwinding from a more brilliant nucleus. This appearance, whose reality had at first been suspect, was fully confirmed — and for ever-increasing numbers of nebulae — when photography came to be applied systematically to their study, about the middle of the century. Exposures of several hours register much more of the delicate structure of the spiral nebulae than can be detected visually.

The earliest serious photographic observations of the brighter spirals were made by the English amateur astronomer Isaac Roberts, in about 1890, with a telescope of 20 inches aperture (Fig. 677 inset), and some ten years later by the Americans G. W. Ritchey and J. E. Keeler, with, respectively, the Yerkes 24-inch and the Lick 36-inch. Impressive collections of photographs have been accumulated since that time by, in particular, Ritchey with the Mount Wilson 60-inch reflector, J. C. Duncan and E. Hubble with the Mount Wilson 100-inch, and H. D. Curtiss with the Lick 36-inch. In more recent times, de Kérolyr has obtained some beautiful plates with the 31-inch telescope at the astrophysical station of the Paris Observatory in Haute Provence.

This mass of material — covering several hundred of the brightest nebulae and some thousands of the fainter ones — enabled Hubble to classify the extragalactic nebulae into three main categories:

1. *Globular and elliqtical nebulae.* With even the largest instruments, these remain completely structureless — mere patches of nebulosity whose brightness descreases steadily from the centre outwards. Their various shapes form a series of increasingly flattened ellipses, from perfectly circular objects like NGC 3379 to lenticular objects like NGC 3115. Their degree of flattening is denoted by indices from 0 to 7, which are a measure of the relative lengths of the two axes: thus NGC 3379, which is perfectly circular, is classed as E0, a slightly flattened nebula like NGC 221 (M 32) is E2, NGC 4621 (M 59), still more flattened, is E5, and finally the lenticular nebulae such as NGC 3115 are classed as E7. It is noticeable that there are no elliptical nebulae more flattened than E7, which suggests that this represents a limit beyond which a nebula in rapid rotation (as their appearance as well as spectroscopic measures show them to be) becomes unstable. We shall return later to this point, which is relevant to the evoluion of the galaxies.

* The volume of space that can be explored by the Mount Palomar 200-inch reflector has a *radius* of about 10^3 light years.

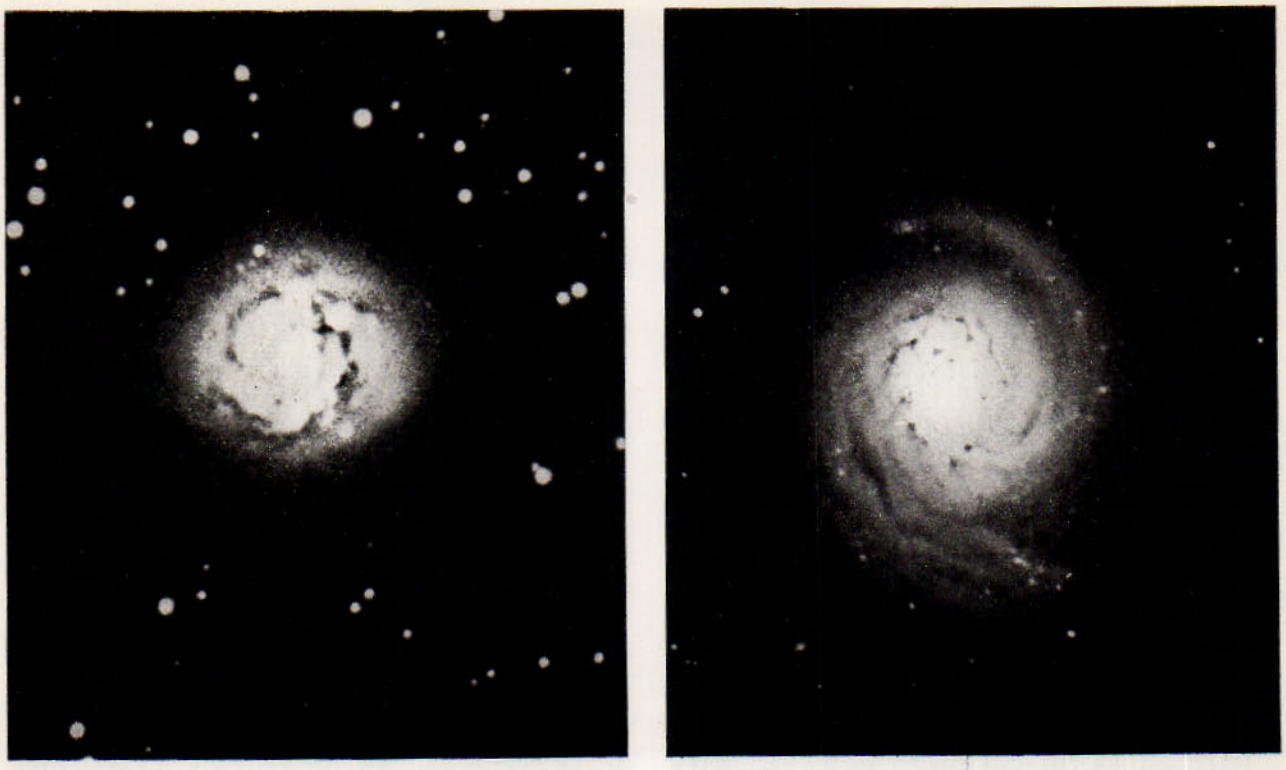


Fig. 684 — TWO NORMAL SPIRALS (TYPE *Sa*) SEEN IN PLAN. *Left:* NGC 278 in Cassiopeia *(photo, Mount Wilson Observatory, 60-inch telescope, exposure 4h; 1912); right:* NGC 4736 in Canes Venatici *(same source, exposure 3h 30m; 1912).*

It must also be realised that the observed oblateness of a nebula will in general be different from its actual (and unknown) degree of flattening, since the spatial orientations of the extragalactic nebulae are quite haphazard: an apparently circular nebula, for example, could in reality be either a sphere or a more or less flattened one seen in plan. In fact, the apparent flattening of a nebula depends both on its real oblateness and on its orientation relative to the observer, and it is impossible to distinguish between them. Only in E7 nebulae can we be sure that we are viewing them edge-on and that the flattening is real, since it is the greatest degree of flattening that occurs.

Among 600 of the brightest nebulae classified by Hubble, 17% were of the elliptical type; of these, the quasi-spherical Eo and E1 nebulae were the commonest. But when the relative abundance of different degrees of real flattening are calculated, it is found that the very flattened E6 and E7 nebulae are in reality commoner than the spherical ones.

Fig. 685 — TWO BARRED SPIRALS, SEEN APPROXIMATELY IN PLAN. *Left:* H. I. 84 in Coma Berenices, type SB*b*; *right:* NGC 7479 in Pegasus, type SB*c*. *(Photos, Lick Observatory, 36-inch telescope; 1902, 1908)*

Fig. 686 — BARRED SPIRAL (NGC 1300) INTERMEDIATE IN TYPE BETWEEN SB*b* AND SB*c*.
(Photo, G. van Biesbroeck, MacDonald Observatory, 80-inch telescope)

2. *Spiral nebulae.* These account for at least 80% of the extragalactic nebulae that have been catalogued. They fall into two distinct sub-types:

(*a*) *Normal spirals*, which make up about 50% of the total. They are divided among three classes — denoted by S*a*, S*b* and S*c* — according to the relative size of the nucleus and the degree of unwinding and development of the arms. S*a* spirals have very large nuclei, around which lie regular and relatively undeveloped arms. When seen edge-on (e.g. NGC 4594) they resemble the planet Saturn, encircled by its rings; an equatorial band of absorbing matter in the plane of the arms even suggests the shadow of the rings on the body of the planet. S*b* spirals (like M 81 in Ursa Major or M 31 in Andromeda) have a relatively smaller nucleus and well developed arms which are more or less condensed into individual knots and lumps, and which often make a complete circle about the nucleus. Finally, in S*c* spirals (e.g. M 101 in Ursa Major and M 33 in Triangulum) the nucleus is very small compared with the spiral arms, which may be either slender (e.g. M 101) or massive (e.g. M 33), but which are always of complex and irregular structure, unwound widely from the nucleus.

When nebulae of this type are seen in plan (i.e. from a position on the prolongation of the minor axis, at right angles to the edge-on position) they generally exhibit two principal arms, which issue tangentially from diametrically opposed points on the edge of the nucleus.

b) *Barred spirals*, which are denoted by SB and comprise about 30% of the total. They are distinguished from normal spirals by the fact that the spiral arms appear to issue from the two extremities of a more or less lengthy bar of matter issuing *radially* from the nucleus, at two diametrically opposed points. Three sub-types are distinguished: in SB*a* spirals, such as NGC 2859, the bar is at first all that can be seen, appearing as a pair of more or less elongated protrusions on either side of the nucleus, which itself resembles an elliptical nebula; more developed examples show a fairly regular circle or ellipse of nebulous material surrounding the nucleus (which is now, relatively, smaller), crossed diametrically by the bar; the whole structure somewhat resembles a Greek ⊕. In SB*b* spirals (e.g. NGC 5850) the ring is broken at the ends of the bar, and thus assumes the form of two arms which issue from the extremities of the bar and at right angles to it, and thence wind tightly round the nucleus. Finally, in SB*c* spirals, such as NGC 7479, the wide-open arms are nothing more than prolongations of the bar, and hence appear to emerge normally from the nucleus at two diametrically opposed points.

3. *Irregular nebulae.* These are less numerous than the members of the first two types, comprising only 2% to 3% of the extragalactic nebulae. About half of them are objects like the Magellanic Clouds (NGC 4449 and NGC 6822 are examples) which are characterised by the absence of a nucleus and a general lack of symmetry; these constitute a homogeneous sub-type. The rest are abnormal nebulae of all sorts, which frankly defy classification; examples are the structureless M 82 in Ursa Major, the one-armed spiral NGC 4039, and NGC 5128 which resembles the nucleus of a spiral surrounded by a whorl of absorbing matter, but which has no arms. All such objects are at present, and lacking any more precise classification, lumped together as 'pathological' cases.

Nebulae of the first two types — elliptical nebulae, and normal and barred spirals — which together make up 97% to 98% of the extragalactic total, can, according to Hubble, be arranged in a continuous sequence which splits into two branches after E7; the transition from E0 to E7, from S*a* to S*c* and from SB*a* to SB*c*, is perfectly smooth, whereas that from E7 to S*a* and SB*a* is more sudden. There are, in fact, very few examples of nebulae intermediate in type between E7 and S*a* or SB*a*; such as there are belong to the sub-type S0. On the other hand, there are a number of objects which appear to be transitional between normal and barred spirals. In short this classification should be regarded as no more than a

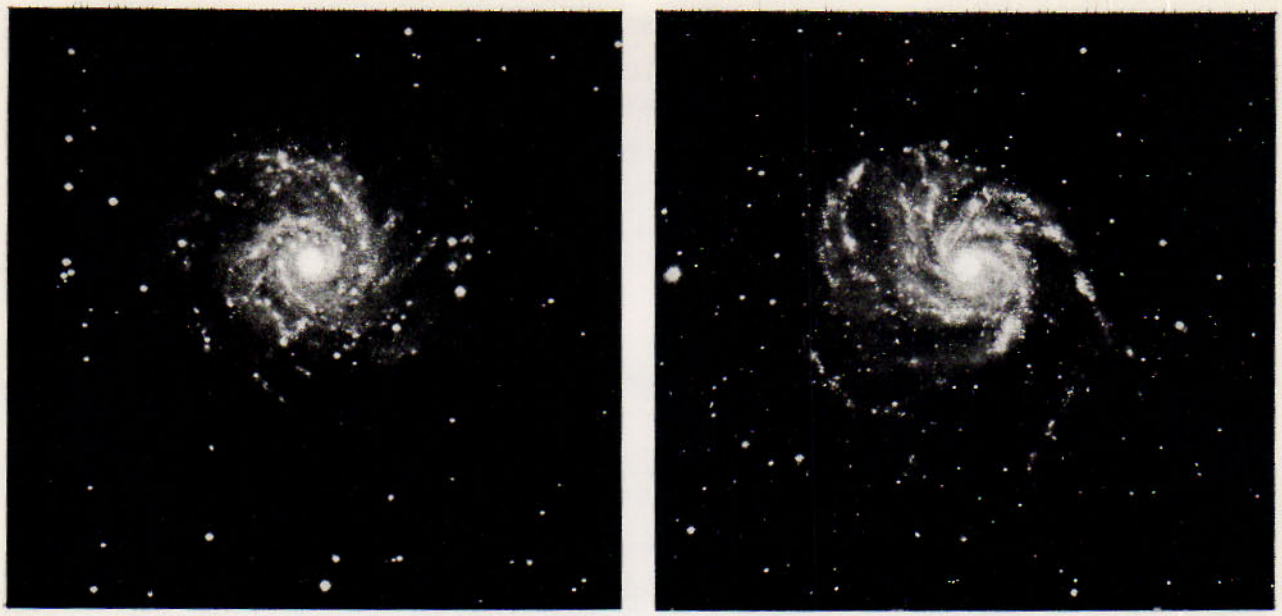


Fig. 687 — TWO NORMAL SPIRALS (TYPE Sc) WITH SLENDER ARMS. *Left*, M 74 in Pisces; *right*, M 101 in Ursa Major. *(Photos, Lick Observatory, 36-inch telescope)*

convenient method of establishing a provisional order among the multifarious nebular types, and there is no proof that the series actually represents the evolutionary sequence followed by each nebula. Furthermore, the abnormal nebulae, and the irregular Magellanic-type nebulae, are very difficult to fit into the scheme, although, an attractive modern hypothesis invokes collission between galaxies as a possible mechanism.

General characteristics of the extragalactic nebulae. — This purely morphological classification of the extragalactic nebulae also co-ordinates a considerable number of their general characteristics, about which something must now be said.

It was first of all established that the apparent diameters of the nebulae of any given apparent magnitude — as determined from photographs taken with large telescopes — develop progressively from one end of the series to the other. Thus, taking Hubble's figures, the angular diameters of magnitude 10 nebulae increase from about 2′ at type Eo to 5′ at E7, after which they range from 6′ to 9′ along the spiral branch from *Sa* to *Sc*. This variation suggests that — as had already been argued by Jeans — spiral arms are ejected from the nuclei of nebulae when they reach a certain stage in their evolution, corresponding to the transitional type SO. According to these figures, the nuclei of the spirals appear to be comparable in size with the elliptical nebulae.

This hypothesis has in fact been to a large extent abandoned, since accurate photometric measurements of

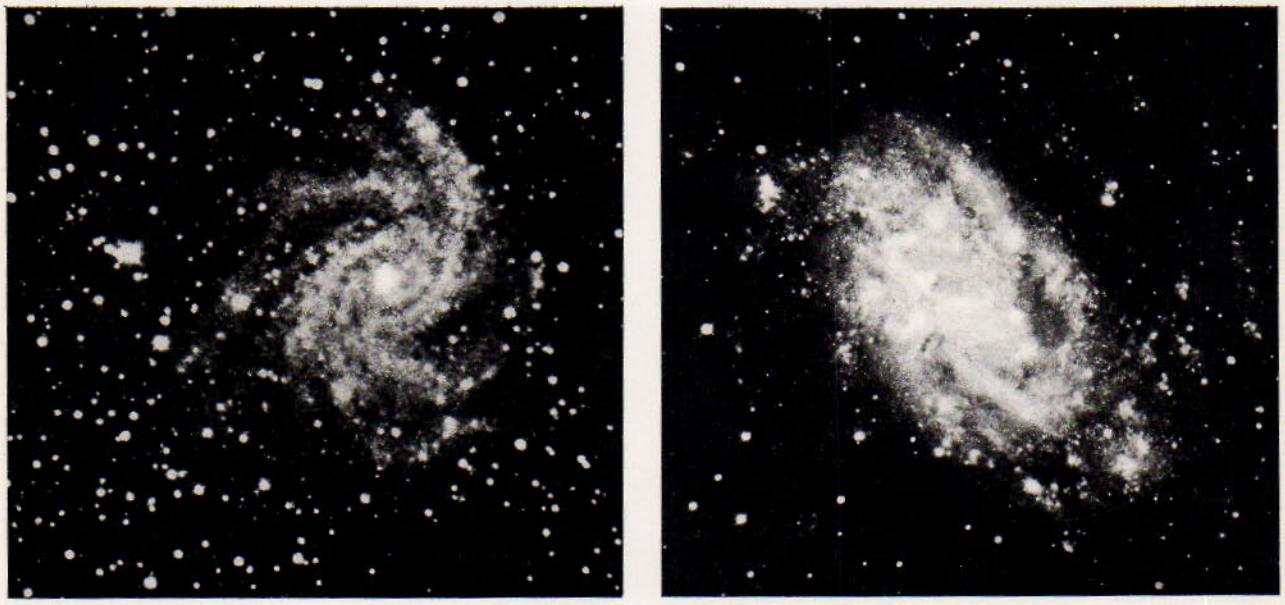


Fig. 688 — TWO NORMAL SPIRALS (TYPE Sc) WITH MASSIVE ARMS. *Left:* NGC 6946 in Cepheus *(photo, Yerkes Observatory, 24-inch telescope, exposure 1h; 1933); right:* NGC 2403 in Camelopardus *(photo, Mount Wilson Observatory, 60-inch telescope, exposure 3h 30m; 1912)*. Note the enormous obscuring cloud to the left of the nucleus of NGC 2403.

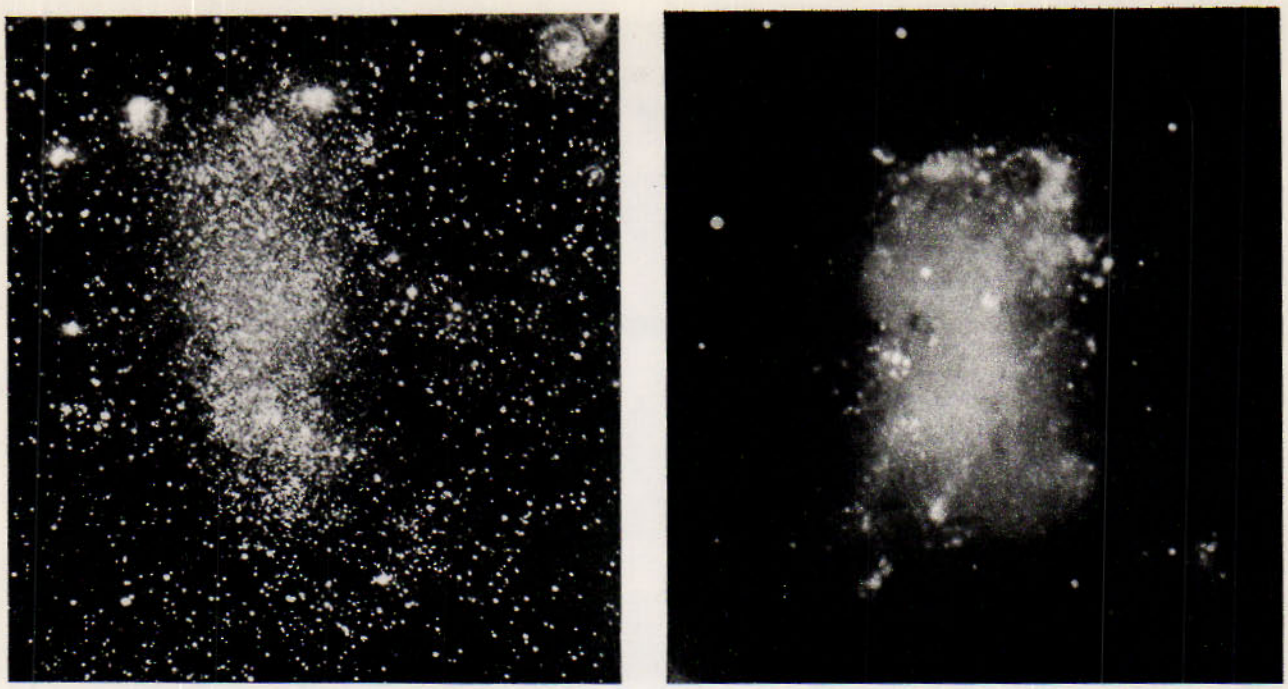


Fig. 689 — TWO IRREGULAR GALAXIES OF THE SAME TYPE AS THE MAGAELLANIC CLOUDS. *Left:* NGC 6822 in Sagittarius, a member of the Local Group. It is resolved into stars in this photograph taken with the Mount Wilson 100-inch reflector, exposure 3h 30m, 10 July 1923. The two irregular patches above it are gaseous nebulae connected with the star cloud. *Right:* NGC 4449 in Canes Venatici, more distant than the preceding, shows no stellar resolution save for the most luminous supergiants. *(Same source, 60-inch reflector, exposure 5h; 7 April 1913)*

photographic negatives have, within the last 25 years, shown that these nebulae generally, and the elliptical nebulae in particular, extend with rapidly decreasing brightness over much wider areas than the ordinary visual inspection of the plates would lead one to believe. This shows that the maximum dimensions of the different types of extragalactic nebula are on the whole much the same, and that their apparent differences in size spring simply from the fact that the arms of the spirals are easier to see than the outer regions of the elliptical nebulae, which fade by imperceptible degrees.

These researches — undertaken, in particular, by Hubble at Mount Wilson, and more recently by R. O. Redman at Cambridge — show that it is in any case impossible to attribute hard-and-fast boundaries and diameters to the extragalactic nebulae. This is especially true of the elliptical nebulae, which appear to have no sharply defined 'edge' at all: measurements made further and further from the nucleus show that the brightness falls off steadily, until finally it is lost in the general brightness of the background of the night sky. The genesis of this diffused luminosity is discussed in Chapter 11,

Fig. 690 — M 102 (NGC 5866) IN BOÖTES APPEARS TO BE AN SB*a* BARRED SPIRAL SEEN EDGE-ON, WITH A FINE THREAD OF ABSORBING MATTER CROSSING ITS NUCLEUS. *(Photo, Mount Wilson Observatory, 60-inch reflector, exposure 2h 40m; 1912*

Fig. 691 — THE ELLIPTICAL NEBULA NGC 3115 IN SEXTANS, TYPE E7, SEEN EDGE-ON; IT SHOWS NO TRACE OF PERIPHERAL ABSORBING MATTER. *(Photo, Mount Wilson Observatory, 60-inch telescope, exposure 1h 40m; 1911)*

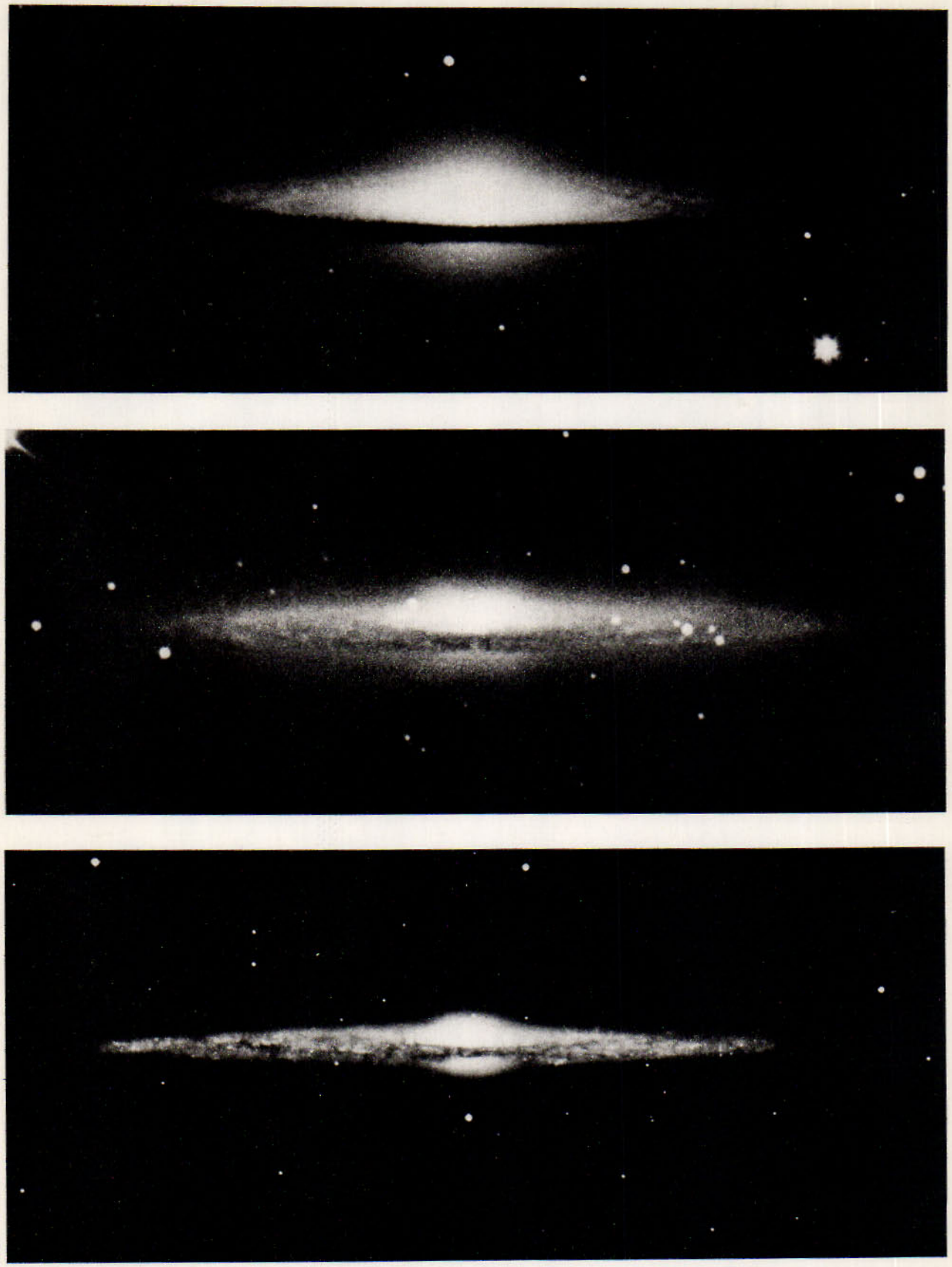


Fig. 692 — THREE FINE NORMAL SPIRALS VIEWED ALMOST EDGE-ON. *Top:* NGC 4594 in Sextans, type S*a*, exposure 2h 15m; 1914. Centre: NGC 5746 in Virgo, type S*b*, exposure 6h; 1914. *Bottom:* NGC 4565 in Coma Berenices, type S*c*, exposure 7h 15m; 1910. Note the progressive development of the arms at the expense of the nucleus.
(Photos, Mount Wilson Observatory, 60-inch telescope)

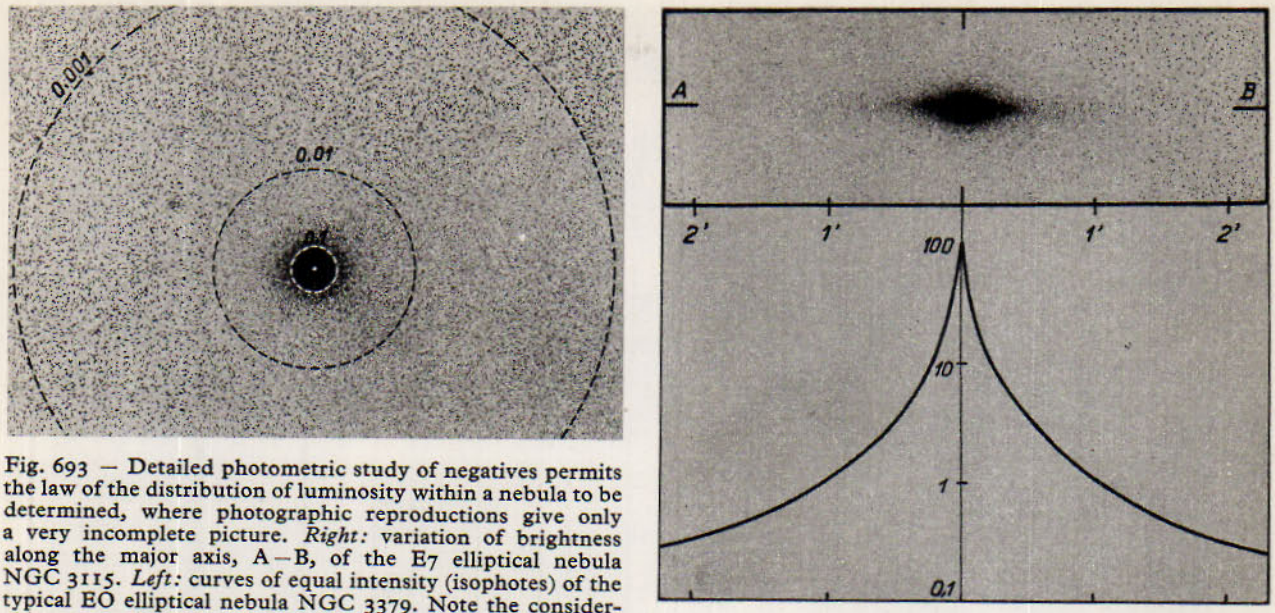


Fig. 693 — Detailed photometric study of negatives permits the law of the distribution of luminosity within a nebula to be determined, where photographic reproductions give only a very incomplete picture. *Right:* variation of brightness along the major axis, A—B, of the E7 elliptical nebula NGC 3115. *Left:* curves of equal intensity (isophotes) of the typical EO elliptical nebula NGC 3379. Note the considerable extent of the low-luminosity regions, as compared with the part directly visible on the negative. *(G. de Vaucouleurs, Haute-Provence Observatory, 31-inch telescope; 1946)*

and it is worth mentioning that it is a serious obstacle to extragalactic studies, since to a greater or lesser extent it blots out the faintest nebulae, just as the brightness of the day-time sky obliterates the stars and turns the Moon into a pale simulacrum of its night-time self.

These same measures of the brightness distribution in the photographic images of the extragalactic nebulae have also shown that the space between the arms of the spirals is far from being as dark as it appears at first sight, and that their brightness is in fact an appreciable fraction of that of the arms themselves. The latter stand out so noticeably largely as a result of contrast, in much the same way that the bright coronal streamers seem to be larger and brighter than they really are; photometric measurements prove that both the coronal streamers and the arms of the spirals are much fainter, relatively, than they appear to be.

Two-colour photometric measurements also demonstrate that the arms of a spiral are much richer in blue light and weaker in red light than the integrated light of the nebula. The investigation of the nearest nebulae, which can be resolved into stars by large telescopes, has shown that this is due to the presence of blue giants and supergiants (types O and B) and their associated gaseous nebulae in the spiral arms.

This raises the question of the stellar resolution of the extragalactic nebulae, and the manner in which it varies along the sequence of nebular types. Numerous condensations in the arms of the nearer spirals can easily be detected in the photographs taken with relatively small instruments, and they were noticed in the earliest days of extragalactic photography by Roberts, Ritchey and others. But it was only in 1924 that Hubble, with the great Mount Wilson reflectors, was able to obtain photographs showing the images of individual stars; these were in fact supergiants — at such distances, the only stars bright enough to register on the plates. The brightest of them were identified as blue supergiants of types O and B.

Save among the nearest galaxies, these are in general the only types of star that can be detected, but they are by no means uniformly distributed among the nebulae: they are found scattered throughout the arms of the *Sc* spirals as far as the vicinity of the nucleus, but only in the outermost arms of the *Sf* spirals, and they are altogether absent from those of type *Sa*. None of the 'earlier' type nebulae — and, in particular, no elliptical nebula — appears to be resolvable into high-luminosity stars.

This does not mean that the elliptical nebulae and the nuclei of spirals are not composed of stars, but only that they contain no blue supergiants. Indeed, W. Baade in 1943 succeeded in obtaining photographs in red light, with the Mount Wilson 100-inch, which resolved the central regions of the Andromeda nebula and several nearby elliptical nebulae into teeming clouds of stars of the 21st magnitude; these objects had shown no trace of resolution in earlier photographs of the ordinary kind, taken in blue light.

The reason underlying this preferential distribution of different stellar types among the various categories of nebula, or even in different regions of the same nebula, is not known. But it is certainly of great significance from the viewpoint of cosmogony. In this connexion. Oort suggested in 1946 that the absence of supergiants of high luminosity and therefore great mass (mass-luminosity relation) from elliptical nebulae might ultimately be connected with the absence of interstellar matter in these systems, for we shall learn in the next chapter that diffuse matter now appears to be a necessary condition of the existence of these stars.

It has already been mentioned that strong bands of absorbing material (in every way comparable with that scattered throughout our own Galaxy) are almost always to be observed in spiral nebulae that are favourably oriented — i. e. with their central plane not too highly inclined to the line of sight. The photographs reproduced here show several fine examples of such dark masses — sometimes very regular (as in NGC 5494, an *Sa* spiral), sometimes distributed more irregularly (as in NGC 891 (*Sb*), NGC 4565 (*Sc*), and doubtless in our own Galaxy also). Now, no trace of an equatorial obscuring band has been detected in any elliptical nebula, even among those of type E7 which are seen edge-on (as NGC 3115). However, it cannot be concluded from this that the elliptical nebulae contain no interstellar matter, since the latter's absorbent properties depend to a very large extent on the

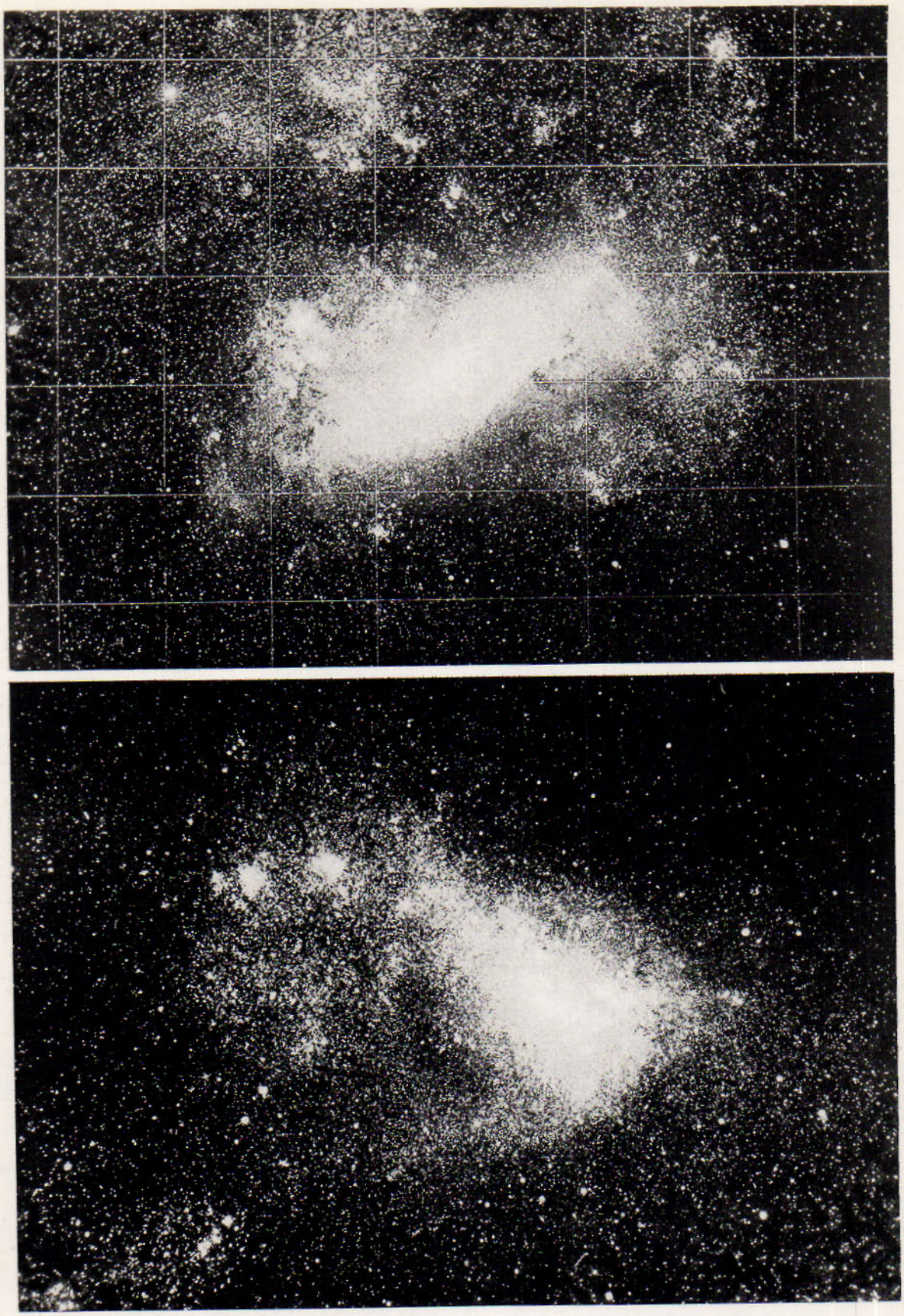


Fig. 694 — *(above)* THE GREATER MAGELLANIC CLOUD. *(Photo, Cape Observatory)*
Fig. 695 — *(below)* THE LESSER MAGELLANIC CLOUD. *(Photo, Arequipa Station of Harvard Observatory)*
The Magellanic Clouds are irregular satellites of our Galaxy. *(Collection, Palais de la Découverte, Paris)*

Fig. 696 — GENERAL VIEW OF M 33 IN TRIANGULUM, TYPE Sc. *(Photo, Mayall, Lick Observatory, 36-inch telescope; 1938)*

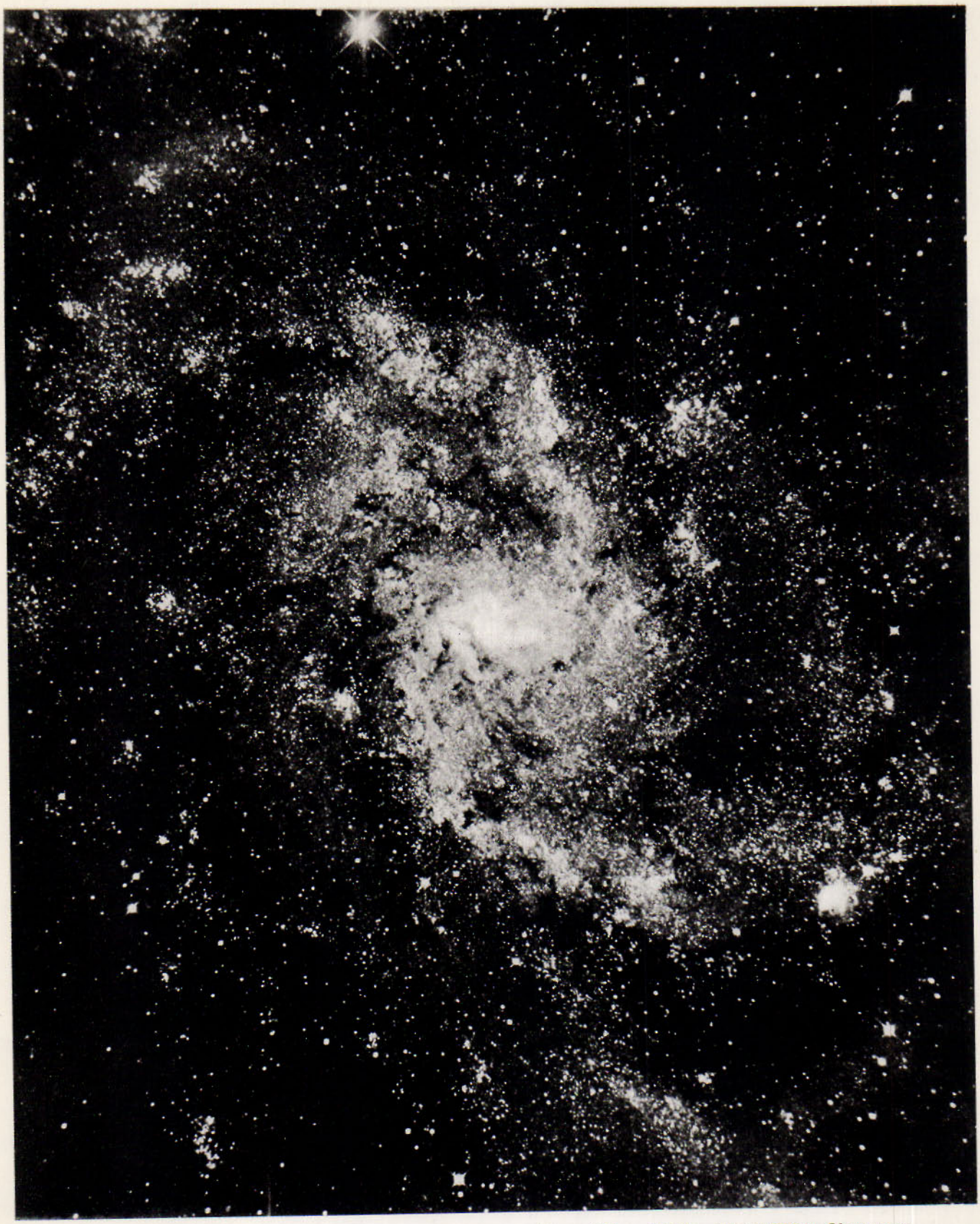


Fig. 697 — STELLAR RESOLUTION OF THE CENTRAL REGION OF M 33. *(Photo, Ritchey, Mount Wilson Observatory, 60-inch telescope, exposure 8h 30m spread over 3 nights; 1910)*

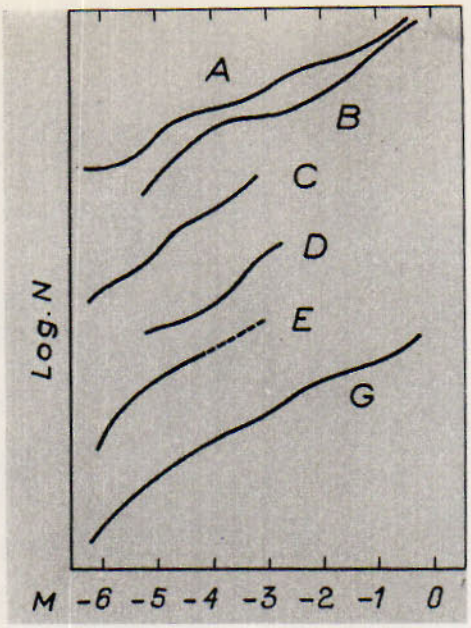


Fig. 698 — COMPARISON OF THE LUMINOSITY FUNCTIONS OF SUPERGIANTS IN THE GALAXY AND IN VARIOUS EXTRAGALACTIC SYSTEMS. A, Greater Magellanic Cloud; B. Lesser Magellanic Cloud; C, the spiral M 33 in Triangulum; D, NGC 6822 (Magellanic type); E, the Andromeda spiral, M 31; G, the Galaxy. *(Seyfert and Nassau, 1946)*

size of its particles. Oort, nevertheless, thinks that this may be the case, and may be the reason for the absence of supergiants in these systems.

Thus elliptical nebulae and spiral nuclei appear to be composed exclusively of stars belonging to the Main Sequence and the red-giant branch of the Hertzsprung-Russell diagram (excluding the supergiants of both), whose absolute visual magnitudes hardly exceed −2; those of the blue supergiants encountered in the arms of 'late' type spirals, on the other hand, are as high as −6 or −7.

This particular Hertzsprung-Russell distribution of the stellar populations of the extragalactic nebulae, though quite different from that encountered in the Galaxy, is matched by that observed in the globular clusters.

It seems certain that all these facts must be of great significance to cosmogony, but what exactly this significance may be, we do not yet know. However that may be, the abundant occurrence of Main Sequence stars too faint to be individually accessible, with the available instruments* is attested by the integrated spectra of the extragalactic nebulae. The spectroscopic types of those nebulae which are bright enough to yield sufficiently detailed spectra are all, in fact, very close to *d*G3, thus revealing the preponderance of solar-type dwarfs. Taken as a whole, the spectra of the extragalactic nebulae (or, rather, of their brighter central regions) range from types A to K, with a marked preference for F and G. Moreover, Hubble and Humason have shown that the mean type changes slightly but progressively along the nebular sequence: G4 for elliptical nebulae, G3 for S*a* and SB*a* spirals, G2 for S*b* and SB*b* spirals, and F9 for S*c* and SB*c* spirals. These are entirely comparable with the mean spectroscopic type of the star clouds in the Milky Way (about F5).

* The absolute magnitude of the faintest stars that can be photographed in the nearest nebulae is still no higher than about 0 — i. e. these stars are about 100 times more luminous than the Sun.

It may be pointed out in passing that all these results confirm the classification of the Galaxy as an S*b* or S*c* spiral, and likewise the Sun's eccentric position within it.

Finally, it must be mentioned that the photoelectric measurements made at Mount Wilson by Stebbins and Whitford have shown that the mean colour index of the extragalactic nebulae appears to be slightly too high for their spectral type, except in the case of the S*c* spirals where the presence of large numbers of blue supergiants more than redresses the balance. This colour excess (amounting to 0.2 or 0.3 magnitude) is difficult to account for in any other way. The effect of selective interstellar absorption has been appealed to in the case of those spirals where the existence of such absorption is established; but this explanation cannot be extended to the elliptical nebulae, which provide no direct evidence of containing any diffuse absorbing material whatever.

Distances and dimensions of the extragalactic nebulae. — Between 1922 and 1924 the extragalactic status of the 'white' nebulae was established, and from that date it became possible to determine their distances, and therefore their real sizes.

The distances of the Magellanic Clouds (70,000 to 80,000 lightyears) were in fact already known at that time, thanks to the Cepheid investigations of Miss Leavitt some ten years previously. But the extragalactic nature of these clouds of stars, and their analogy with the unresolved irregular (extragalactic) nebulae had not yet been recognised.

It was not until Hubble's discovery of variables, identifiable as Cepheids, in the spirals M 31 and M 33, in addition to the Magellanic NGC 6822, that the distances of these objects could be determined beyond doubt by means of the period-luminosity relationship (see Chapter 13). It transpired that they were of the order of one million lightyears, which far exceeded the assigned dimensions of the Galaxy, and immediately established their extragalactic status. They were, in fact, stellar systems external to the Galaxy and comparable with it in size: the 'island universe' hypothesis had been finally vindicated.

This conclusion is nowadays supported by a considerable mass of associated evidence. In the first place, the results of the Cepheid observations cannot be questioned, since in each of the nebulae studied the number of variables was large enough to establish the period-luminosity relation firmly, and it tallied perfectly with that furnished by the Cepheids in the Magellanic Clouds and the globular clusters. Next, at this distance of 700,000 to 750,000 lightyears, given by the Cepheids, the numerous novae* observed in the Andromeda nebula would have absolute magnitudes close to − 6, which makes them altogether comparable with those of our Galaxy — a conclusion which is independently suggested by their spectra and light curves. Again, if we accept this distance, then the numerous globular clusters (about 300) which are observed as small, hazy condensations, a few seconds of arc in diameter, in all regions of the Andromeda spiral, as well as in its immediate neighbourhood, will have absolute magnitudes of from −4 to −7; the same as that of the globular clusters in the Magellanic Clouds, though slightly lower than that of the galactic globular clusters. Finally, star counts to successive magnitude levels within the Andromeda nebula yield the same luminosity function (see p. 336) as that derived for the Galaxy and the Magellanic Clouds, assuming that the Cepheid-derived distance of the nebula is correct (Fig. 638).

Thus there is no room for doubt regarding the general correctness of the distances of the nearest extragalactic

* Well over 100 have been observed during the last 40 years.

systems as derived from observations of their Cepheids. But such observations can, unfortunately, be made in no more than about a dozen of the nearest extragalactic nebulae. If we are to reach out any further into extragalactic space we must make use of other distance criteria, which, though individually less accurate, are applicable at enormously greater distances.

The first of these additional distance criteria to be employed by Hubble had already been applied by Shapley in solving the problem of the distances of the globular clusters — that of the brightest stars in each system. Observations of the nearest extragalactic systems (the Magellanic Clouds, M 31, M 33, NGC 6822 and IC1613), whose distances are given by their Cepheids, had shown that the mean absolute photographic magnitude of the three or four brightest stars (type O supergiants) is about −6 — as in the Galaxy. Once this was established, it was only necessary to determine the apparent magnitudes of a few of the brightest stars in any galaxy, in order to deduce its distance; and the result would usually be accurate to within 20% or 25%. The margin of error in Cepheid distance determinations is only 5% to 10%; but whereas the latter are observable in no more than a handful of the nearest nebulae (distances less than about 10^6 lightyears), the Mount Wilson reflectors are capable of showing the brightest stars in over 100 nebulae, the most distant of which are nearly 10^7 lightyears from the Sun.

It so happens that one of the great clusters of extragalactic nebulae (see p. 419) lies within this distance: the Virgo cluster of several hundred nebulae of all types, about 25% of which are elliptical and 75% spirals. The mean apparent magnitude of the brightest stars in those of type *Sc* is 20.5, equal to a distance of 7 or 8 million lightyears (about 2.5 megaparsecs). But the apparent magnitudes of the nebulae themselves range from about 10.5 to 15.5, which at this distance is equivalent to absolute magnitudes of from −16 to −11.5, or, in the mean, −14. In other words, the total luminosity of a galaxy is in general of the order of 10^7 to 10^8 times that of the Sun, and exceeds that of the brightest known stars by about 8 magnitudes.

The magnitudes of the actual nebulae may thus be used as a new distance criterion — still less precise than the last in individual cases, but applicable to the extreme limit at which extragalactic objects are visible at all.

Highly accurate results cannot be derived by this criterion owing to the wide spread of absolute luminosity among the extragalactic nebulae. A nebula of the 15th magnitude (apparent) would be situated at a distance of 5 million lightyears if it were a dwarf system of absolute magnitude −11 (e. g. NGC 6822), and at a distance of 80 million lightyears if it were a giant like the great Andromeda spiral, with an absolute magnitude of −17. But these extreme cases are fairly rare, and it may be accepted that the mean error in the derived distances does not generally exceed ± 50%; and this error will be considerably reduced when the mean of a number of independent determinations can be taken — as is possible, for example, with the members of a cluster of galaxies.

Thus when, for example, the apparent magnitudes of the nebulae in a cluster range from 15 to 21, with a mean value of 18, the distance of the cluster may be derived as about 80 million lightyears (25 megaparsecs), with a reasonable margin of uncertainty amounting to only 10% or 20%.

This may be the place to point out that allowance has to be made for an effect of statistical selection, due to the fact that among the nebulae of a given apparent magnitude there is a preponderance of giants over dwarfs (the former being visible to much greater distances than the latter); the mean absolute magnitude for nebulae selected according to their apparent magnitude should therefore be taken as −15 instead of −14. This value is applicable to all types of elliptical and spiral nebulae, the absolute magnitudes of which do not appear to vary systematically along the sequence of nebular types.

Fig. 699 — THE EXTENT OF THE ANDROMEDA NEBULA AS REVEALED BY MICROPHOTOMETRIC MEASUREMENTS, COMPARED WITH THE REGION DIRECTLY VISIBLE ON PHOTOGRAPHS. The cross marks the distance from the nucleus that corresponds to the position of the Sun in the Galaxy. *(Williams and Hiltner, 1941)*

Since this uncertainty does not enter into the consideration of clusters of nebulae, where all types are represented, these provide the most reliable values of great distances: they even provide a final distance criterion, analogous to that involving the brightest stars in individual galaxies or globular clusters. It is a fact of observation that the brightest nebulae in each cluster (or the fifth brightest, for example, to eliminate any of abnormal luminosity) have very much the same absolute magnitude (about −16.5). Granted this, one has only to determine the apparent magnitude of the fifth brightest nebula in a remote cluster, or to take the mean of the apparent magnitudes of the half dozen brightest, in order to derive the cluster's distance with a still very reasonable degree of accuracy — to within 20%, perhaps. Thus the distance of a cluster whose fifth brightest member is magnitude 19.5 would in theory be 5×10^8 lightyears (150 megaparsecs). In fact, at such enormous distances, certain corrections have to be applied to the derived figure, as we shall see presently.

It can thus be seen that a highly ingenious series of extrapolations — from Cepheids to the individual brightest stars, from these to individual nebulae, and from these to clusters of nebulae — has enabled distance determinations to be pushed outward step by step until finally they extend to the furthermost limits of the observable Universe. This astonishing achievement was realised in little more than 10 years (1924−34) and was almost wholly the work of one man, Hubble, using the 60-inch and 100-inch reflectors at Mount Wilson.

Knowledge of the distances of the extragalactic nebulae allows their linear dimensions to be directly calculated from their angular dimensions. In this way Hubble has shown that the mean diameters of the brightest central regions of the nebulae (which coincide more or less with their photographic images) are of the order of 2,000

Fig. 700 — SPIRAL NEBULA M 81 IN URSA MAJOR, TYPE Sb. *(Photo, G. W. Ritchey, Mount Wilson Observatory, 60-inch telescope, exposure 3h; 1917)*

Fig. 701 — SPIRAL NEBULA NGC 6946 IN CEPHEUS, TYPE *Sc*. Situated close to the zone of avoidance, it is strongly obscured by the considerable thickness of interstellar matter through which we see it. *(Photo, Mount Wilson Observatory, 100-inch reflector, exposure 4h; 1922)*

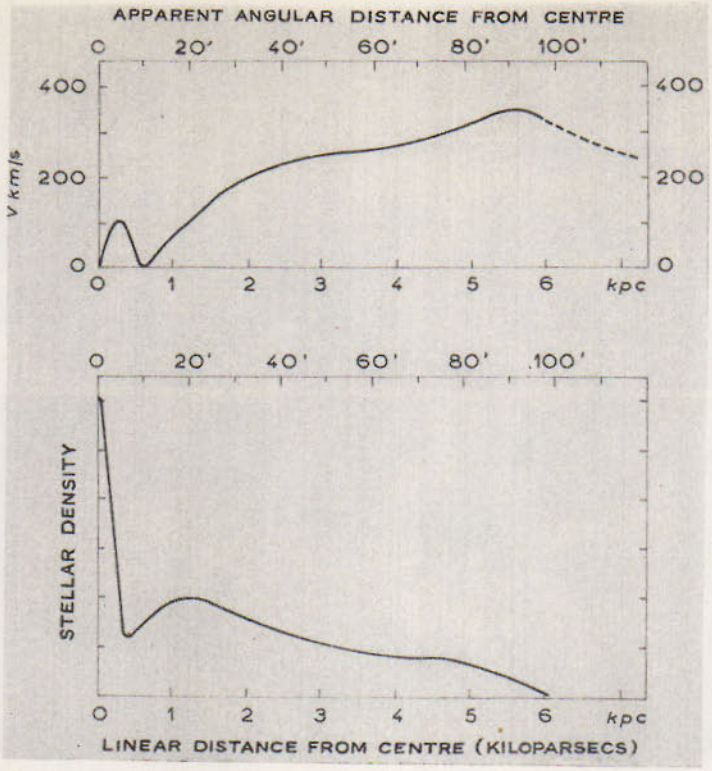


Fig. 702 — DISTRIBUTION OF ROTATIONAL VELOCITIES *(above)* AND MASSES *(below)* IN THE ANDROMEDA SPIRAL, M 31, AS FUNCTIONS OF DISTANCE FROM THE CENTRE. *(Babcock, Wyse and Mayall, Lick Observatory, 1942)*

lightyears for globular nebulae, 4,000—5,000 lightyears for 'late' elliptical nebulae, and 6,000—9,000 in the case of irregular nebulae and spirals. For 20 years or so these values appeared to invalidate the supposed identity of status of the Galaxy and the extragalactic nebulae, for Shapley had attributed a diameter of some 300,000 lightyears to the former; thus, as someone remarked, 'if the extragalactic nebulae are island universes, the Galaxy is a continent'. In an attempt to resolve this antinomy, Shapley even suggested that the Galaxy is in reality not a single star system but a cluster of galaxies, or 'supergalaxy'. But then the difficulty is merely turned inside out, for the Galaxy is much too small to constitute a cluster analogous to those which we actually observe in the heavens.

But here again, we have been gradually forced to recognise that the apparent uniqueness of our home is an illusion: the discrepancy between the derived sizes of the Galaxy and the other nebulae has been largely whittled away, if not entirely removed, and the true and more rational order of things been established. On the one hand, the understanding of the important rôle played by interstellar absorption (reduction of the apparent brightness of distant objects in the Galaxy, thus exaggerating their distances as derived by photometric methods) has led to a considerable reduction of the accepted dimensions of our stellar system; as we saw in the last chapter, its diameter is nowadays accepted as not exceeding about 100,000 lightyears, and many astronomers believe it to be even smaller, of the order of 80,000 lightyears. And on the other hand, photometric measures of the extragalactic nebulae have shown that they extend to much greater distances around their nuclei than a first glance at the photographs would suggest.

The Andromeda nebula has been particularly well investigated in this connexion. The major axis of its photographic image does not exceed 2.5° (150′), or about 35,000 lightyears. But microphotometric examination of the negatives has shown that in facts it extends over a much wider area, with a major axis of at least 4.5° (270′), or nearly 60,000 lightyears. Finally, the system of globular clusters associated with the nebula extends still further, having a diameter of the order of 80,000 lightyears, which is strictly comparable with that of the Galaxy.

These observations allow us to form a clearer appreciation of the Sun's eccentric position in the Galaxy: if it were transposed to the Andromeda spiral and placed at the same distance from the centre (25,000 to 30,000 lightyears) as it is from the centre of the Galaxy, it would be situated some 2° from the nucleus, well outside the limits of the nebula as directly visible in photographs (Fig. 699). As for the dwarf galaxies, such as the companion of M 31, their dimensions appear not to exceed a few thousand lightyears. These systems must nevertheless be classed as extragalactic nebulae rather than globular clusters, since the maximum observed diameters of the latter are to be measured in hundreds of lightyears only. Furthermore, a considerable number of globular clusters have been discovered in association with the nearest dwarf galaxies, which they surround like satellites.

Rotation and masses of the extragalactic nebulae. — The appearance of the spirals — great cosmic whirlpools — and their high degree of flattening are enough to suggest that these gigantic systems are, like the Galaxy, in rapid rotation.

As will be explained later (Chapter 20) this rotation can be directly demonstrated by the inclination of the spectral lines in the case of edge-on nebulae (such as NGC 3115, E7, or NGC 4594, S*a*) when the spectroscope

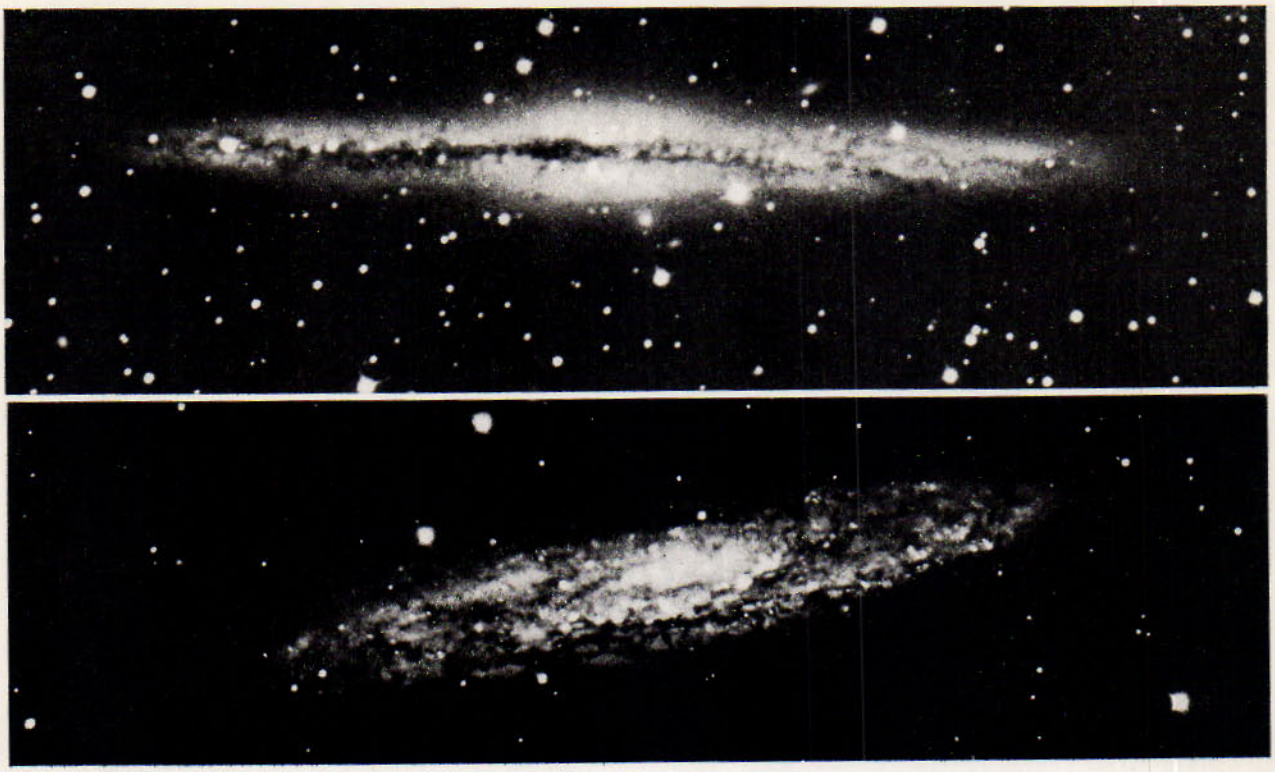


Fig. 703 — Two spiral nebulae, one seen exactly edge-on, the other slightly inclined to the line of sight, showing a complex band of peripheral absorbing matter which indicates, unambiguously, the side nearest the Earth. *Above:* NGC 891 in Andromeda *(Mount Wilson Observatory, 60-inch telescope, exposure 7h 15m; 1916); below:* NGC 253 in Cetus *(Lick Observatory, 36-inch telescope; 1907).*

slit is laid along the major axis of the nebula's image. The radial velocities revealed in this way are of the order of several hundred m.p.s. at different points in the nucleus, and hence are quite comparable with those that have been derived for the Galaxy. These measures indicated that the linear rotational velocity of the nebulae concerned (M 31, NGC 3115, NGC 4595) increases in proportion to distance from the centre of the nucleus: in other words, they appear to rotate all in one piece, like solid bodies. But this result only applies to the immediate vicinity of the nucleus, which is the only region bright enough to yield usable spectra.

Within the last 20 years, moreover, the rotations of the two great spirals, M 31 in Andromeda and M 33 in Triangulum, have been studied in detail at Lick Observatory by Babcock, Wyse, Mayall and Aller. By observing the emission-line spectra of the gaseous nebulae which occur at scattered points throughout these spirals, they were able to derive the distribution of radial velocities, and from them to deduce the rotational velocities, relative to the nucleus, at different distances from the nucleus.

From these results it appears that M 31 and M 33 rotate in an entirely different manner from that described above: in M 33 the rotational velocity increases regularly from the nucleus outward until it reaches a value of from 60 to 100 m.p.s. at a distance of 1 kiloparsec (3,200 lightyears) from the nucleus; thereafter it decreases slowly till its value is halved at about 2 kiloparsecs (6,500 lightyears) from the centre; beyond this, observations could not be made. In M 31 the velocity of rotation increases rapidly at first, to 60 m.p.s. at 200 parsecs from the centre, then decreases equally rapidly to zero at about 500 parsecs (1600 light years); at this distance, which is still very near to the nucleus (4′), the nebula does not appear to rotate at all. From here outward the rotational velocity increases again, attaining values of 60 m.p.s. at 1,200 parsecs (4,000 lightyears) from the centre, 120 m.p.s. at 2 kiloparsecs (6,500 lightyears), 200 m.p.s. at a little over 4 kiloparsecs (13,000 lightyears), and finally 220 m.p.s. at rather less than 6 kiloparsecs (18,000 lightyears) — i. e. towards the edge of the region that can be distinctly seen in photographs; beyond this it seems likely that it begins to decrease slowly.

From these observed velocity distributions the probable distribution of mass within the nebulae can be calculated theoretically, and it has been found that the density of M 33 is virtually constant throughout the whole of the region lying within about 1 kiloparsec (17′) of the centre. Between 1 and 2 kiloparsecs from the centre the mass is negligible, corresponding at most to 20% of the total mass of the nebula, which is estimated as 2×10^9 times that of the Sun. The density of the central region works out as 2 solar masses per cubic parsec, and that of the surrounding ring about one-tenth of this. The latter figure compares very closely with that of the Galaxy in the vicinity of the Sun, which Oort puts at not more than 0.1 solar masses per cubic parsec (including both stars and interstellar matter). This once again testifies to the very eccentric position of the Sun within the Galaxy.

These detailed data regarding the distribution of mass in the large galaxies comparable with our own — the first to be obtained — are of the greatest value in understanding the dynamics of the Galaxy itself, though this introduces other complications which we cannot go into here.

Another important problem which is engaging the attention of astronomers at the present time is the direction of rotation of the spirals: do they rotate in the direction of the unwinding of the arms, or against it? This question is important, since it is tied up with that of the origin and mode of formation of the spiral arms. And according

Fig. 704 — (a) *Left:* spiral nebula in Pegasus (NGC 7331), studied by Hubble and Lindblad, which present an even more clearly marked absorption band; it is, however, inclined at too great an angle to the line of sight for us to be sure whether or not it is situated on the side nearest the Earth — i. e. whether it is peripheral or internal. *(Photo, J. E. Keeler, Lick Observatory, 36-inch telescope; 1899)*
(4) *Right:* the band of obscuring material near the nucleus of this spiral nebula in Coma Berenices (M 64) is obviously situated between two arms of the spiral and is not peripheral. *(Photo, Lick Observatory, 36-inch telescope; 1900)*

to the answer that is given, theories regarding this further question will be either confirmed or contradicted.

Forty years ago it was thought that the answer could be found quite simply by measuring very accurately the positions of condensations in the arms of spirals, relative to galactic stars whose images lie near them on the photographic plate, and then to repeat the measurement after a sufficiently long lapse of time; and it was believed that infinitesimal rotational motions had indeed been detected in a number of spiral nebulae. But it later became evident that these measures were affected by complex systematic errors which completely invalidated the results, and more recent work has failed to confirm them. Furthermore, it is now realised that the rotations of the spirals are much too slow for their effects to be measurable on photographs taken at an interval or only a score or so of years.

Fig. 705 — THIS EDGE-ON SPIRAL NEBULA IN PEGASUS (NCG 7814) BEARS A CURIOUS RESEMBLANCE TO AN ELECTRIC ARC BETWEEN TWO POINTED ELECTRODES. *(Photo, J. E. Keeler, Lick Observatory, 36-inch telescope; 1899)*

For example, the observed rotational velocity of 60 m.p.s. at a distance of 1 kiloparsec from the centre of M 33 corresponds to a rotation period of about 50 to 70 million years, and an angular displacement of only one thousandth of a second of arc per century. This is far below the present-day threshold of accuracy in the measurement of photographic plates, which is about 0.001 mm. — corresponding, at the best, to only 0.01″ in the focal plane of the largest telescopes. With our present equipment and observational methods we should therefore have to wait for at least a thousand years before the rotation of the nearest spirals could produce measurable displacements on the photographic plates.

This being so, attempts have been made to solve the problem of the direction of rotation of the spirals by means of the radial velocities of a number of edge-on nebulae — or, at any rate, of nebulae whose planes are so little inclined to the line of sight that the radial component of their rotational velocity is of appreciable size. But here a fundamental difficulty is encountered. It is perfectly true that the spectroscopic measurements indicate quite clearly which end of the major axis of a nebula is approaching, and which receding; but in order to discover its direction of rotation we need to know more than this — we must know its actual orientation on space, so that we can distinguish its nearer edge from that which lies on the far side of the nucleus.

Now there is in fact something that provides this very information — the absorbing band which is projected against the nucleus of edge-on spirals. But if a nebula is exactly edge-on, its spiral structure cannot be seen, and the problem is indeterminate. On the other hand, when the plane of the nebula is sufficiently inclined to the line of sight for the arrangement of the arms to be seen clearly, the absorbing band is displaced by perspective above or below the nucleus, and now it cannot be seen; thus the spatial orientation of the nebula still evades us. However, as a result of a systematic search of the Mount Wilson files, Hubble in 1942 succeeded in finding several photographs of nebulae whose approaching and receding sides had been determined spectroscopically, and whose inclination to the line of sight was just great enough for

Fig. 706 — A PAIR OF SPIRALS (NGC 4567, 4568) IN VIRGO. *(Photo, F. G. Pease, Mount Wilson Observatoty, 60-inch telescope, exposure 6h; 1914)*

the layout of the arms to be seen, but not so great that the peripheral absorbing band, indicating the near side, was invisible.

By means of these he was able to demonstrate that all the spiral nebulae rotate in the same sense, and that this is such that the arms trail *behind* the nucleus as the whole rotates; in other words, the direction of rotation is such as would tend to wind up the arms.

This is the conclusion that was reached by V.M.Slipher at Lowell Observatory, some 40 years ago, but, unhappily, not that of the Swedish authority, Bertil Lindblad, who at the present time rejects it completely. He maintains that the direction of rotation is such as tends to unwind the arms, which thus *precede* the nucleus. He was led to this conclusion in the first place by theoretical consideration based on his investigations of the formation of spirals, but equally he appeals to observation for support of his views. He contends that in spirals which are sufficiently inclined for the layout of their arms to be seen, the zone of strongest absorption does not indicate the near side but the side furthest from the observer, since this zone does not lie outside the outermost spiral arm (as Hubble thought), but between it and the arm next inside it. Lindblad furthermore believes that he has been able to establish, by detailed photometric analysis, that the side further from the observer is redder than the other, and since a colour excess of this sort is encountered in regions of absorption he concludes that the direction of rotation is the opposite of that advocated by Hubble.

The problems thus remains unresolved. It may, moreover, be less simple than was originally thought, for Lindblad has drawn attention to the case of a nebula whose outer arms unwind in the opposite direction to the inner, and within the Galaxy itself evidence is accumulating that some stars revolve in the opposite direction to the majority.

Groups, clusters and clouds of nebulae. — The small-scale distribution of the extragalactic nebulae is visibly not random: employing the current nomenclature for increasingly large associations of nebulae, they show a marked propensity to congregate in pairs, triplets or multiplets, then in larger groups and widely extended clusters, and finally in gigantic clouds. Even allowing for accidental associations due to perspective (analogous to optical doubles as distinct from binary stars) it is clear that a considerable proportion of the galaxies belong to systems of two, three or more members, even outside the true clusters (which we will come to shortly).

This tendency is illustrated by the following table, which quotes E. Holmberg's statistics for the relative frequency of isolated nebulae ($n = 1$) and nebulae which are members of double or multiple systems of increasing size ($n = 2, 3 \ldots$), among 400 non-cluster nebulae whose apparent magnitudes lie between 10 and 13:

FREQUENCY OF EXTRAGALACTIC NEBULAE BELONGING TO MULTIPLE SYSTEMS

NUMBER OF NEBULAE IN SYSTEM	PERCENTAGE OF TOTAL
1	47
2	24
3	15
4	7
5	4
6	(2)
7	(1)

Thus less than 50% of the extragalactic nebulae appear to be isolated in space; more often than not, they belong to a group (large or small) of nebulae.

Our Galaxy is itself a member of such a group — known as the Local Group (not to be confused with the Local Cluster of stars in the Sun's vicinity) — the number of whose firmly established members is now 13. These, together with some relevant data, are listed below:

GALAXIES BELONGING TO THE LOCAL GROUP (W. Baade)

GALAXY	TYPE	DISTANCE		PHOTOGRAPHIC MAGNITUDE		DIAMETER		
							LINEAR	
		KILO PARSECS	10^3 LIGHT-YEARS	APPARENT	ABSOLUTE	APPARENT	KILO PARSECS	10^3 LIGHT YEARS
The Galaxy	*Sb*?	—	—	—	— 18 ?	—	24	78
M 31 Andromedae	*Sb*	231	750	5±	— 17.9	3.2°	12.9	42
Nubecula Major	I	22	70	0.5	— 15.9	12°	4.6	15
M 33 Trianguli	*Sc*	239	780	7.8	— 14.9	62′	4.3	14
Nubecula Minor	I	25	80	1.5±	— 14.5	8°	3.6	12
M 32 Andromedae	E2	231	750	9.5	— 12.9	—	—	—
Fornax system	E*p*	142	460	9.0	— 11.9	50′	2.1	6.8
NGC 205 Andromedae	E5*p*	231	750	10.8	— 11.5	15.8′	1.1	3.6
NGC 6822 Sagittarii ..	I	161	515	11.0	— 10.8	20′	0.94	3.0
IC 1613	I	225	730	—	— 10.8	17′	1.1	3.6
Sculptor system	E*p*	69	225	9.0	— 10.6	45′	0.90	2.9
NGC 185 Andromedae	E	204±	665±	11.8	— 10.6	14.5′	0.86	2.8
NGC 147 Andromedae	E	204±	665±	12.1	— 10.3	14.1′	0.83	2.7
Possible members: IC 10 (*Sc*?), IC 342 (*Sc*), NGC 6946 (*Sc*)								

The status of the three doubtful members is uncertain because of their proximity to the zone of avoidance, where interstellar absorption falsifies the distance determinations, and the necessary corrections are not yet known accurately enough for the derived distances to be accepted as satisfactory.

The volume of space occupied by the Local Group is, roughly, a flattened ellipsoid, with a major axis of about 10^6 lightyears, a minor axis about half this, and a thickness of from 200,000 to 300,000 lightyears. The centre lies in the general direction of the Andromeda nebula at a distance of some 500,000 lightyears from the Galaxy, which, together with its two satellites, the Magellanic Clouds, is situated at one extremity of the group.

The Local Group seems to differ in several important respects from other, similar associations of galaxies scattered throughout space. Its members include a couple of spirals (the Galaxy and M 31) which must rank as giants in the nebular hierarchy, quite a number of dwarf systems (NGC 147, NGC 185, NGC 205, and the Sculptor and Fornax systems), and also an abnormal number of irregular nebulae (IC 1613, NGC 6822 and the Magellanic Clouds), two of which are dwarf systems. Furthermore, the Group includes no *Sa* or barred spiral. The only normal members, typical of those found outside the Group, are M 32, M 33 and the Magellanic Clouds.

One wonders whether the Local Group is indeed an exception to the general run of similar associations of galaxies, or whether (as seems more likely, especially as regards its relative abundance of dwarf galaxies) our existing information about the constitution of distant groups is still incomplete — just as in the Galaxy the millions of dwarf stars are invisible, save those in the Sun's vicinity. At the present time it appears that the absolute magnitudes of the galaxies range from about −10 (extreme dwarf systems) to −18 (the largest systems, such as M 31): a luminosity range of more than 10,000:1, whereas their diameters only vary from about 1 to 20 or 30 kiloparsecs. These values, incidentally, show that even the smallest elliptical nebulae are still considerably larger than the largest globular clusters, whose diameters do not exceed 100 parsecs, although their stellar density is a score or more times greater.

The table on p. 418 shows that the frequency of groups of nebulae falls off rapidly as the number of members in the group increases; hence the number of clusters of galaxies, properly so called, is comparatively very small. The cluster in Virgo is (in angular measure) the largest — because the nearest — and also the one that has been most thoroughly investigated. Its central region (measuring more than 20° across) covers a large part of the constellations of Coma and Virgo, and at its distance of 7 to 8 million lightyears this corresponds to a major axis of 2 to 3 million lightyears. Although its brightest members (about magnitude 10) have been studied in detail, its total population is still undetermined, since it is difficult to distinguish between dwarfs belonging to the cluster, and more distant and apparently faint 'background' nebulae. Its overall dimensions may, moreover, be considerably larger than those of the richest region, quoted above. An abnormally large number of bright nebulae are in fact to be observed on either side of it, and also on the prolongation of its major axis, in the constellations Ursa Major and Lynx to the north, and Virgo and Centaurus to the south; these, if included, would give the cluster an extent of at least 60°, corresponding to a major axis of over 8 million lightyears. It has even been suggested that the Local Group may be a secondary condensation on the outskirts of a great cloud of nebulae, of which the Virgo cluster constitutes the most highly condensed central region; but this question is still undecided.

There is, however, no doubt that the Virgo cluster is a vast supergalactic organisation of several hundred nebulae into a system whose shape is visibly flattened. Because of this flattening, attempts have been made by means of the analysis of the radial velocities of the member galaxies to discover whether the cluster as a whole may not be rotating. The investigations of S. Smith at Mount Wilson, some 25 years ago, seemed to indicate that this is not the case. But discussion of the distribution of the individual radial velocities within the cluster suggested that the escape velocity is of the order of 1,000 m.p.s., which in turn indicated a total mass in the region of 10^{14} times that of the Sun; in this case the mean individual nebular mass would be about 2×10^{11} that of the Sun. Although this value agrees well with that derived for the Galaxy, it seems to be much too high (10 to 100 times) for a typical extragalactic nebula, as explained above.

F. Zwicky's studies of the Coma cluster at Mount Palomar (1942) have shown that the spatial distribution of the members of this cluster (as of several others that he investigated) is analogous to that of the molecules in a gaseous sphere in gravitational and isothermal equilibrium. Now, the mean mass of the particles (molecules or galaxies) in such a system can be related to their mean individual random motions within the system, and on this basis the Mount Wilson observations indicate a mean random velocity of the order of 1,300 m.p.s. in the Coma cluster; from this the mean mass of the component galaxies may be deduced as 2×10^{11} times that of the Sun.

Since masses of the same order were also derived by Zwicky by other methods, we are faced with a discrepancy, whose solution is still being sought. The most plausible explanation yet proposed envisages the existence of great quantities of non-luminous internebular matter within the cluster — but so far no direct observational evidence of its existence has been obtained.

A number of other clusters, of various sizes, have been observed. The principal data for the dozen that have been studied in sufficient detail are listed in the table below:

THE PRINCIPAL CLUSTERS OF GALAXIES

CONSTELLATION	APPARENT DIAMETER (nucleus)	DISTANCE (10^6 light-years)	RADIAL VELOCITY		MAGNITUDE OF BRIGHTEST NEBULAE
			km/sec.	m.p.s.	
Virgo No. 1	12°	7.5	1,200	700	10.5
Pegasus	1°	28	3,800	2,400	13
Cancer	1.5°	29	4,800	3,000	13.5
Perseus	2°	35	5,200	3,200	14
Coma	2°	45	7,400	4,600	14.5
Ursa Major No. 1	0.7°	85	15,500	9,300	16
Leo	0.6°	115	19,600	12,400	16.5
Corona Borealis	0.5°	125	21,200	13,000	17
Gemini		135	23,400	14,000	17
Boötes		230	39,000	24,000	18
Ursa Major No. 2		240	42,000	26,000	18
Hydra		360	61,000	38,000	19

Fig. 707 — GROUP OF GALAXIES OF VARIOUS TYPES, NEAR TO NGC 7317—7320 IN PEGASUS; discovered by Stephan at Marseilles Observatory towards the end of the nineteenth century. *(Photo, Mount Wilson Observatory, 60-inch reflector, exposure 7h 45m; 1916)*

Fig. 708 — FIELD OF SMALL SPIRAL AND ELLIPTICAL NEBULAE IN THE VICINITY OF NGC 7782 IN PISCES. *(Photo, Mount Wilson Observatory, 60-inch telescope; exposure 4h 15m; 1917)*

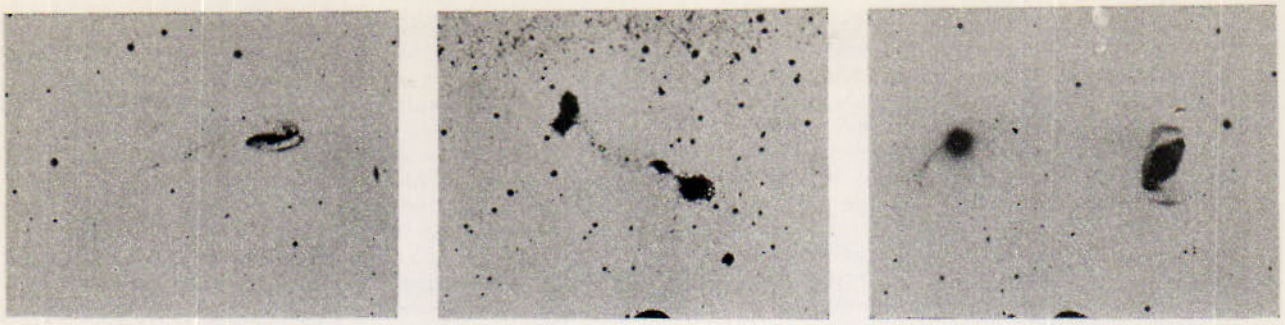


Fig. 709, 710, 711 — INSTANCES OF INTERGALACTIC COSMIC MATTER AND OF 'BRIDGES' OF MATTER CONNECTING NEIGHBOURING GALAXIES

Thus the clusters of nebulae normally contain several hundred members. Their general symmetry and strong central condensation distinguish them clearly from the general field of nebulae. About 25 have been listed within a distance of 200 to 250 million lightyears of the Sun.

The term 'clouds' is reserved for the much more extensive associations of galaxies whose existence was revealed by the counts made at Harvard Observatory about 20 years ago. Among the most important of these clouds may be mentioned that which extends across Andromeda, Triangulum, Pegasus and Pisces, and comprises several tens of thousands of faint nebulae to magnitude 18; its distance is certainly of the order of 10^8 lightyears, and its linear extent several 10^7 lightyears. Another vast cloud of galaxies lies in the direction of Pictor and Dorado at a distance of over 10^8 lightyears, while in Hercules there is a double cloud of faint nebulae.

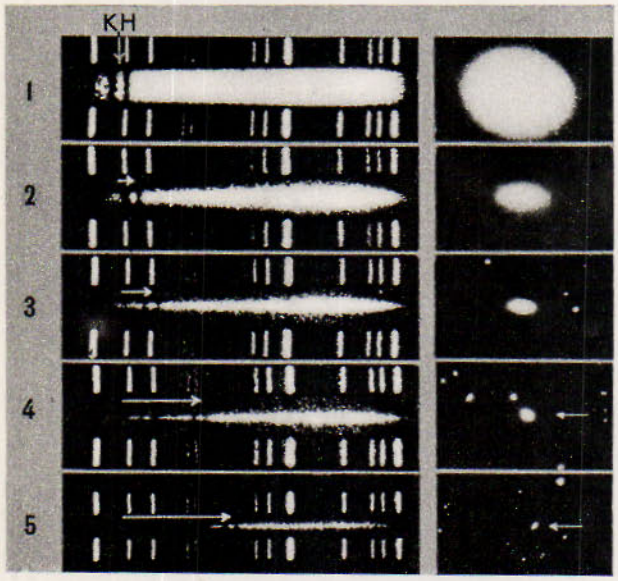


Fig. 712 — RELATION BETWEEN DISTANCE AND RADIAL VELOCITY AMONG EXTRAGALACTIC NEBULAE. The shift of the H and K lines of their spectr *(left)* increases with their distance.

1. NGC 221 (M 32) in Andromeda: radial velocity −120 m. p. s., distance 750,000 lightyears.
2. NGC 4473 in Virgo: velocity +1,400 m. p. s., distance $7 \cdot 10^6$ lightyears (2.5 megaparsecs).
3. NGC 379 in Pisces: velocity +3,400 m. p. s., distance $2.3 \cdot 10^7$ lightyears (7 megaparsecs).
4. Nebula belonging to the Ursa Major cluster (No. 1): velocity +9,400 m. p. s., distance $8.5 \cdot 10^7$ lightyears (26 megaparsecs).
5. Nebula in the Gemini cluster: velocity +14,000 m. p. s., distance $1.35 \cdot 10^8$ lightyears (42 megaparsecs).

The insignificant apparent velocity of approach of M 32 results from the Sun's galactic revolution *(Hubble and Humason, Mount Wilson Observatory, 1935)*

The investigation of these clouds is still in its infancy, but it may be excepted to develop rapidly when the Harvard Catalogue of nebulae brighter than magnitude 18 is completed. We shall then no doubt be able to form a better conception of the general layout of the universe of the galaxies — the Metagalaxy — and, in this, establish a second large-scale property of the visible Universe. For one such property — totally unexpected and still imperfectly understood — has already been discovered.

The existence of cosmic matter in intergalactic space. — Before 1950, absolutely no observational evidence had been obtained of absorption (whether total or differential) that could be attributed to cosmic material in intergalactic space. Such evidence has, however, been provided by the powerful Mount Palomar instrument, and the existence of diffuse intergalactic matter is now universally accepted. At a conference held at the Institut d'Astrophysique de Paris in 1954, under the chairmanship of Professor F. Zwicky, apparently incontrovertible evidence was produced:

(1) Faint blue stars had already been observed in the directions of the poles of our Galaxy. And there is every reason for believing that they lie far beyond the limits of our stellar system — that they are, in fact, intergalactic strays.

(2) As the power of observational instruments increased, so more and more precise information was obtained regarding the outer limits of various galaxies and of their exact apparent diameters. This led to a considerable increase in accepted estimates of these quantities: the major axis of M 31 in Andromeda, for instance, cannot be less than about 15° in length, which means that the two satellite galaxies are enmeshed in its more outlying regions. It is therefore by no means absurd to envisage the possibility of the intermingling of the outermost regions of neighbouring galaxies.

(3) Arms of matter connecting spatially proximate galaxies have, in any case, been detected by direct observation. The list of double and triple systems that are known to be linked in this way is being constantly added to. Some very remarkable photographs, showing such bridges of matter, are reproduced in negative in Figs. 709—711.

These features are most commonly bluish in colour (indicating matter at a high temperature), though some are relatively reddish. Their linear extent may exceed 300,000 parsecs (about 10^6 light years).

Not infrequently a solitary arm is to be seen projecting from a point diametrically opposite to that from which one of these intergalactic bridges springs — a phenomenon that is readily explicable by the general theory of tides.

(4) It is a fact of observation that the central regions of the great clusters of galaxies are suffused with bright diffuse material — though whether this consists of individual stars or dwarf galaxies we do not know.

(5) Lastly, counts of faint galaxies have revealed a falling off of their numbers throughout the area covered by a cluster of larger galaxies: this may well indicate the presence of absorbing material within the clusters of extragalactic nebulae.

The mean density of matter in space is an essential datum in theories of an expanding Universe (pp. 446 to 447). It is still too early to say to what extent the currently accepted value (10^{-31} gm/cm^3.) should be modified in order to take into account the cosmic matter whose existence has been clearly indicated by these observations.

Radial velocities of the galaxies, and the expansion of the Universe. — We have already learnt how the displacement of the lines in the spectrum of a celestial object can be measured by comparing this spectrum with that of a laboratory source, and that from this shift the body's radial velocity can be derived — i.e. its velocity of approach or recession in the line of sight.

When this technique was first applied to the extragalactic nebulae, some 40 years ago, velocities of 1,000 m.p.s. and more were encountered, far surpassing the tens, or at most the few hundreds, of m.p.s. of the stars. But at that time no general law of extragalactic velocities could be deduced, since exact data relating to the distances of the galaxies were lacking.

It was only in 1929 that Hubble, having determined the distances of a sufficient number of galaxies with adequate accuracy, was able to announce that an astonishing relationship links the distances and radial velocities of the extragalactic nebulae: when the Sun's motion due to the rotation of the Galaxy was allowed for, it was found that all the nebulae save two or three of the nearest are receding from us; and, furthermore, the greater the distance of a nebula, the greater its velocity — the radial velocities of recession increasing regularly by about 300 m.p.s. per megaparsec. The nebulae of the Virgo cluster, at a distance of 7×10^6 lightyears, appear to be receding at a mean velocity of 600 m.p.s.

During the following years, from 1930 to 1936, this relationship was confirmed, and its scope extended, by the measures made at Mount Wilson by Hubble and Humason of the radial velocities of a further 150 or so nebulae; these included both isolated specimens in the general field and members of clusters, distributed throughout almost the whole region of space observable at that time. In order to obtain measurable images of the spectra of the remotest and faintest nebulae it was necessary to develop special short-focus (therefore very fast) spectrograph objectives, and to expose the plates for 10, 20, or more hours, spreading the exposure over several nights.

These observations showed that throughout the whole observable region the radial velocities of recession continue to increase steadily with distance, attaining the fantastic value of 35,000 m.p.s. (one-seventh of the speed of light) at a distance of 2.4×10^8 lightyears — that of the remotest nebula for which a spectrogram could be obtained.

It is this extraordinary phenomenon that constitutes the so-called 'expansion of the Universe'. For we cannot believe that the galaxies are really scattering away from us in all directions: such a state of affairs would put the Galaxy back in a unique position in the Universe, and furthermore make it the centre of a universal 'repulsive force' — something so inherently improbable as to be quite incredible. It must rather be supposed that space itself is expanding — a possibility that had already been suggested by the theory of relativity — and in doing so is carrying the galaxies with it. Hence the distances of every galaxy in the Universe from every other galaxy will be continuously increasing, and it would make no difference from which one we happen to make our observations. The result in every case would be the same: an apparent recession of all the others.

A clearer idea of this state of affairs may be given by the following well-known analogy. Suppose that the extragalactic nebulae are represented by droplets on the surface of a soap bubble which is being blown up: then it can easily be seen that as the bubble swells, the distances separating the droplets will increase; furthermore, they will increase with a velocity that is proportional to their separation. No matter which droplet we consider, all the others will appear to be receding from it; hence none has a more unique or central position than another. The 'radial' velocities of the droplets are of course those

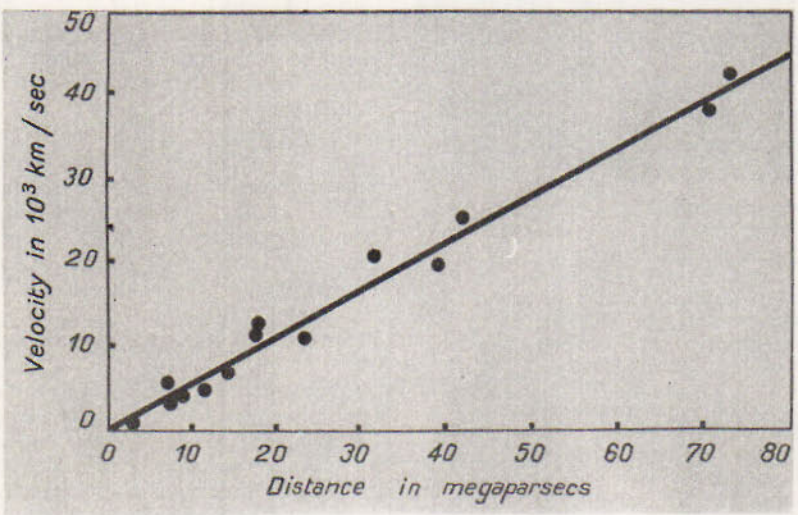


Fig. 713 — THE EXPANSION OF THE UNIVERSE, EXPRESSED AS A LINEAR RELATION BETWEEN THE VELOCITY OF RECESSION AND THE DISTANCE OF THE EXTRAGALACTIC NEBULAE. *(Hubble and Humason)*

of their displacements measured across the surface of the bubble, whose expansion is taking place *in* space, about a centre which is situated in its interior. It is generally agreed that the apparent recession of the galaxies must result from an analogous expansion *of* space — the four-dimensional space-time of relativity theory.

The observed phenomena have in fact been reasonably well accounted for by the theory of an expanding Universe as developed some thirty years ago by the Belgian cosmologist, the Abbé G. Lemaître, as a solution of the fundamental equations for the structure of cosmic space deduced from Einstein's general theory of relativity. A detailed account of Lemaître's theory would take us far beyond the scope of a book such as this, but some

Fig. 714 — THE URSA MAJOR CLUSTER OF NEBULAE (No. 1), DISTANT SOME 75 TO 85 MILION LIGHTYEARS. The arrow indicates NGC 4884, the brightest member of the cluster, which is receding at a velocity of 7,000 m.p.s.
(Photo, Mount Wilson Observatory, 60-inch telescope; 1930)

account of the genesis of the theory may nevertheless be attempted.

The theory of relativity led Einstein to the conclusion that the Universe as a whole constitutes a homogeneous, four-dimensional continuum (three dimensions of space, and one of time), which is finite yet unbounded. The analogy with a sphere, such as the Earth, may help to clarify this conception, for although the Earth's dimensions are finite we can nevertheless move for an infinite distance in any direction over its surface without coming to any limiting 'edge'.

Now the radius of Einstein's hypersphere is related to a quantity known as the cosmological constant, whose value depends on the mean density of matter and radiation in space. Einstein had originally determined the radius that corresponded to the observed density of matter, but had taken into account neither its motion nor the radiation in the Universe. A more thoroughgoing mathematical analysis of the problem showed that such a universe, having the radius calculated by Einstein, would be unstable, for he assumed a certain relation between the cosmological constant and the mean density of matter to have been verified with rigorous precision; now, to assume that this relation held good at some given moment in time is not to say that it still does so in the physical Universe of today, owing to the ceaseless motion of its material content. The instability of Einstein's model, which neglects the motion of matter, means that its radius must vary as a function of time — the model must be either expanding or contracting.

Another solution, worked out in 1917 by the Dutch astronomer, de Sitter, also led to a universe whose radius varies as a function of time; but he assumed a nil density, which, likewise, clearly applies to no real case.

A more plausible solution involving a variable radius was that propounded by the Abbé Lemaître in 1927; this took the existence of matter into account but ignored the tensions set up by matter and radiation, which seemed permissible in view of their extreme dilution in the present state of the Universe. Given this, and assuming a different value for the relation between the cosmological constant and the total mass of the Universe*, he derived the elements of a universe intermediate between those of de Sitter and Einstein, which accounted for the observed positive radial velocities of the galaxies. Several 10^9 years ago, according to this hypothesis, the Universe was an Einstein universe with a radius of about 10^9 lightyears; its present-day radius is perhaps ten times as great; and since it doubles its size every 2×10^9 years it is tending without limit towards the 'empty' de Sitter model.

The wide acceptance won by this hypothesis of a Universe in expansion is explained by the fact that it offers a reason for the observed radial velocities of the extragalactic nebulae. Its speculative character is nevertheless obvious, for although its foundation is the general theory of relativity it nevertheless involves a number of arbitrary assumpions.

Various astronomers and mathematicians — among them de Sitter, Tolman and others, and Einstein himself — have attempted to free the theory from these assumptions, so as to obtain absolutely general solutions for the world equations. A whole series of possible universes, expanding or oscillating, have thus been investigated; some with positive curvature (i.e. closed), some with zero curvature (i.e. infinite and Euclidean), and others again with negative curvature (i.e. hyperbolic and infinite). Of these various models, the one which most closely matches the observed Universe must be chosen; according to Hubble's and Tolman's discussion of the problem, this criterion would seem to favour the expanding spherical model — at any rate it disqualifies those with negative curvature.

The observational data — the precise formulation of the law of proportionality between radial velocity and distance, and, above all, the mean density of matter in the Universe — are in fact not yet known with sufficient accuracy for an unambiguous choice to be made between these alternatives. But that is not the only difficulty — or even the most important one. Another, still more serious, is that it is not yet absolutely established that the apparent expansion of the Universe is real. Doubts have been cast on the interpretation of the red-shifts of the nebular spectra in terms of enormous radial velocities. There are several reason for entertaining such a doubt. In the first place, it seems rash to go straight ahead and apply to such distant objects and such enormous shifts a law that has been established only for the relatively near stars and their comparatively insignificant velocities. Then again, the dimensions and density of the spherical model which agrees most closely with observation involve so small a radius that it barely exceeds that of the domain already explored, while its mean density is so high as to be almost incredible. Finally, if the present trend of the expansion of the Universe is extrapolated backward in time, it is found that all the matter of the Universe must have been condensed into an extremely small volume only 2×10^9 years ago, which would seem to be too short a time span to accommodate even the Earth's geological past. This question of the past history and age of the Universe will be more fully discussed in the next chapter, but the first two objections alone are, in Hubble's opinion, weighty enough to justify a radical re-examination of the whole problem.

This problem can be stated as follows: is there any observational means of deciding whether the red-shifts really indicate positive radial velocities, or whether — the nebulae being in fact more or less stationary relative to the observer — they are caused by some other and still unknown phenomenon, dependent on the immense distances of the galaxies?

Hubble and Tolman have shown that there is in fact a criterion whereby these two alternatives may be distinguished, and they have even attempted to apply it. The principle involved is this. A nebula which is really receding from the observer will appear fainter than an identical but stationary nebula at the same distance. The photon density in a beam of radiation from the receding nebula will be lower than that in one from the stationary nebula, since the photons that are emitted in, say, one second will be strung out along a greater length of the light path; this increase will be exactly equal to the proportion between the source's radial velocity and the velocity of light. This is the so-called 'number effect', and it is independent of the wavelength of the radiation. Its value is simple to calculate, and it is found that it amounts to about 0.15 magnitude for very remote nebulae whose velocity of recession is 26,000 m.p.s. The number effect is thus not very large, though it is large enough to be detected.

But unfortunately it is combined with another effect, and this is larger and much less easy to evaluate, and modifies the radiation of both the receding and the stationary nebula equally. According to the fundamental quantum laws (see Chapter 20), the mere fact of the displacement of all radiations towards the red — whatever the cause of the displacement may be — must involve the diminution of the energy carried by each photon; this diminution is proportional to the relative change in wavelength. The total energy of the radiation received

* Or its density, which comes to the same thing since the volume of the Universe (V) is finite. In elliptical (closed) space, $V = P^2 R^3$ and its circumference $= P R$, where R is its radius.

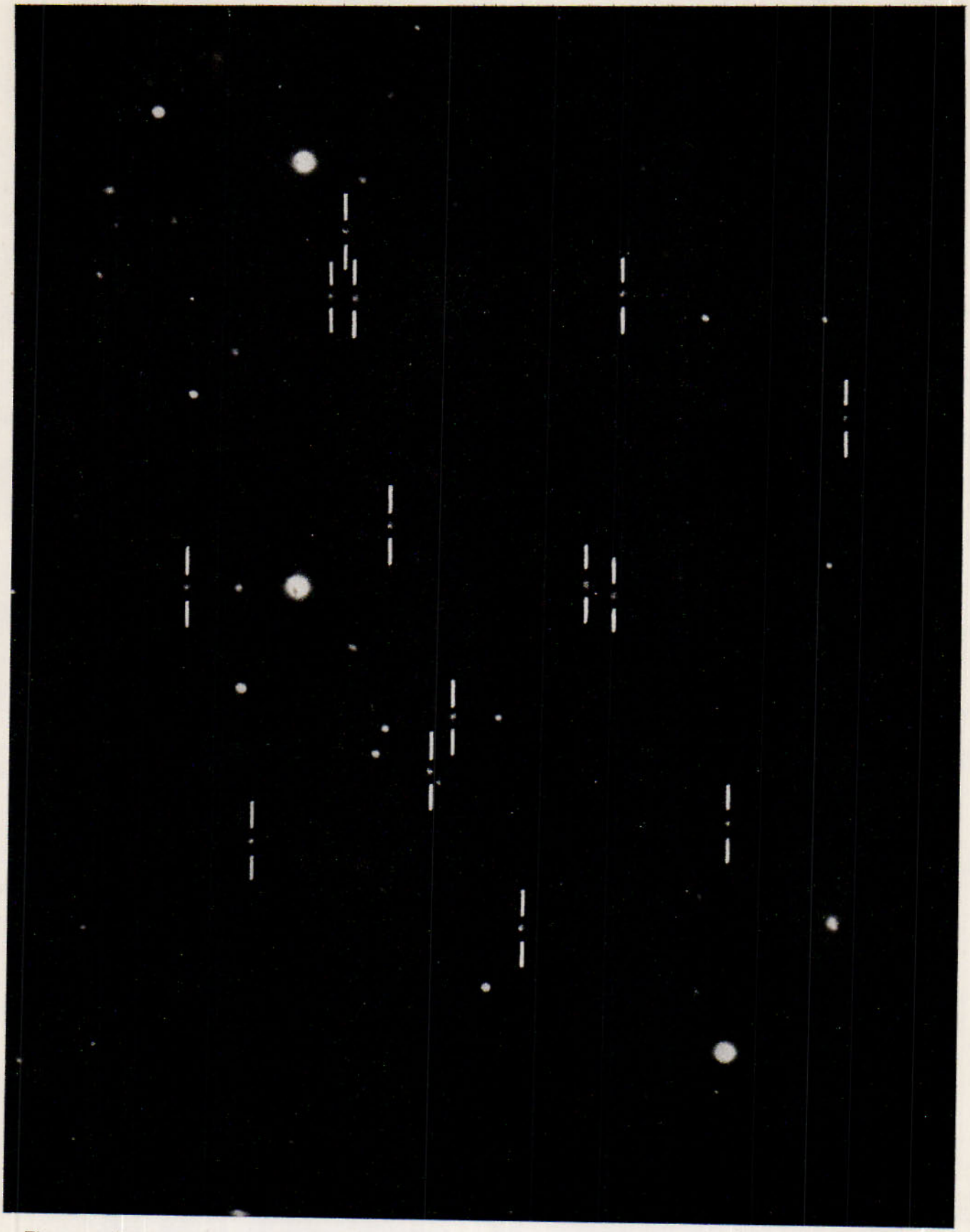


Fig. 715 — A FIELD OF FAINT NEBULAE IN COMA BERENICES. Between each pair of white lines on the original plate from which this picture is reproduced, can be detected a faint smudge — all that can be seen even with the great 200-inch telescope of galaxies at the extreme limit of our observation, 10^9 lightyears away or more. *(Photo, Mount Palomar, 200-inch Hale telescope)*

from a galaxy will therefore be reduced in a proportion which, as in the case of the number effect, is easy to calculate. This is the so-called 'energy effect'. But in fact we do not receive anything like the total radiation emitted by the nebula: the terrestrial atmosphere acts as a filter, the reaction of the photographic emulsion is selective (favouring the short wavelengths), and the telescope mirrors likewise reflect radiations of different wavelengths unequally. A large correction therefore has to be applied to the observational data, and the precise value of this correction is extremely difficult to determine. For this reason the distinguishing criterion proposed by Hubble and Tolman loses much of its value.

Hubble nevertheless attempted to apply it. It involves determining the manner in which the numbers of the fainter galaxies increase with decreasing brightness (and therefore increasing distance) and then comparing this with what would have been anticipated from similar counts of the brightest nebulae which are near enough for the effect of their red-shifts to be negligible. In this way it was established that the increase of numbers with decreasing brightness is slower than would have been expected, and that the discrepancy itself increases at progressively lower limiting magnitudes. Thus in every direction the galaxies appear to thin out at very great distances. Once again the observer seems to find himself in a privileged position in the Universe — which implies that something is wrong. The apparent thinning out of the nebulae at great distances must rather be an effect of their red-shifts, which, as anticipated, make the nebulae appear fainter than they really are. The large-scale homogeneity of the Universe can be restored by applying to the apparent magnitudes a correction which increases from 0 for the near nebulae to 0.3 magnitude at magnitude 19, and about 0.6 magnitude at 21. It is found that the theoretical corrections derived from the hypothesis that the red-shifts are due to some still unknown effect agree more closely with these empirical corrections than those derived from the alternative hypothesis that the red-shifts indicate recession. Hubble was thus led to favour a model which, if not static and infinite, is at least in very slow expansion (or contraction), and whose radius is much greater than that of the observable region; in this model, furthermore, the red-shifts of the nebular spectra are a reflection of some still undetermined physical principle.

However, these conclusions (which have been subject to energetic criticism) entail numerous uncertainties, and no final solution of the problem of the expansion of the Universe is yet in sight. In any case, until further information is available, the majority of astronomers are content to regard the theory of an expanding Universe (such as that developed by Lemaître), in which the red-shifts are the ordinary Doppler effect, as the only one which is consonant both with the observational results and with the known phenomena of physics. Despite all the doubts and difficulties, therefore, it may be accepted with reasonable confidence, but with the reservation that final confirmation must await further observations. Great hopes are held of the results that can be obtained with the 200-inch Mount Palomar instrument, which can reach out 1,000 million lightyears into space and reveal radial velocities of up to about 60,000 m.p.s., or one-third of the speed of light.

All else belongs, at the present time, to the realm of speculation, but if the exact proportionality between recession and distance continues indefinitely, we must envisage a time in the not so distant future when we shall, so to speak, run out of space: we shall find ourselves confronted by an impassable barrier, at a distance of the order of 1.5×10^9 lightyears, where the velocity of recession of the galaxies becomes equal to that of light. No photon from a galaxy at this limit could, in fact, ever reach the observer, and the Universe beyond would necessarily remain forever unobservable. As we near this ultimate barrier we shall be warned of the fact by the progressive thinning out of the remotest nebulae, whose entire spectra would finally be displaced into the infra-red.

But though such speculations are nothing more than extrapolation (possibly over-extended) from the observational data available today, one may justly be proud of the astonishing surge of discovery which, in less than 30 years, has almost at a bound extended our knowledge of the Universe to distances of nearly 10^9 lightyears — the limit of light-grasp of the largest instrument at present in operation.

Armed with these new data on the scale of the Universe, and with recently discovered facts from the domain of the atomic nucleus — so apparently different and yet in reality so closely connected — the theoreticians are applying themselves with renewed energy to the vast and perennial problem of the past and future history of the Universe. The next chapter is devoted to a brief survey of some of the more recent speculations in this field.

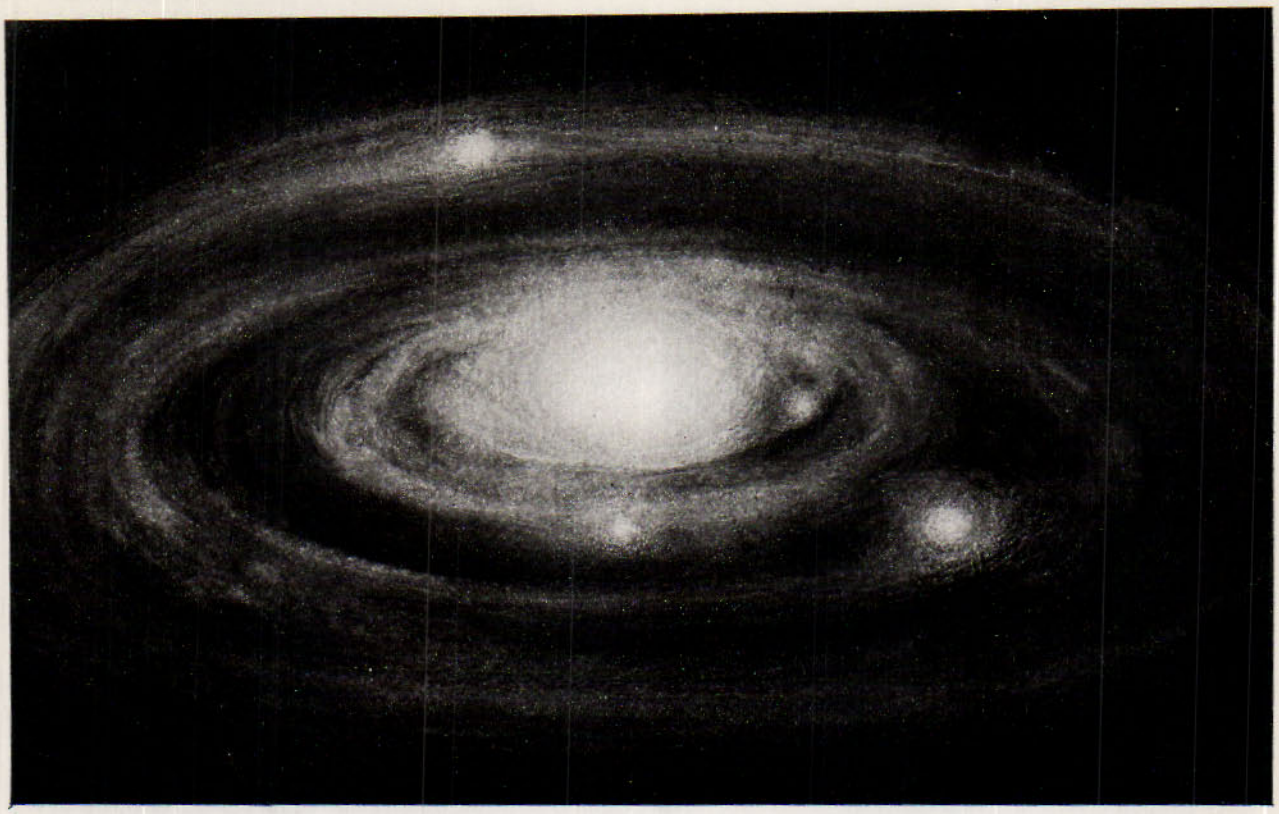


Fig. 716 — LAPLACE'S COSMOGONY, SINCE ABANDONED, SOUGHT TO EXPLAIN THE FORMATION OF THE SUN AND PLANETS IN TERMS OF THE PROGRESSIVE CONTRACTION OF A PRIMAL NEBULA

CHAPTER 18

PAST AND FUTURE

Up to this point we have confined our enquiries to the distribution, motions and nature of the celestial bodies, as revealed by direct observation and by the application of the proven theories of physics; we have now completed the inventory of the heavens and summed up our astronomical and physical knowledge of the present state of the Universe. But still this is not enough: Man's eternal desire is to be able to peer into the future; and now that he has learnt something of the present-day Universe he longs to know where it all came from and where it is going, what was its origin and what its ultimate destiny.

To be a little more precise, we should like to know the age of the Earth as a separate planet, and the method of its formation; the ages of the stars, of the galaxies, and of the Universe as a whole; the way in which all these were 'born', and subsequently organised in their present-day dispositions; and, finally, how the stars evolved, what the future holds for them (if this can be discerned), and the scale of future time still lying ahead of us.

All these questions belong to the most theoretical of all the branches of astronomy — cosmogony. Cosmogony has already produced numerous theories, ranging from the naïve anthropocentrism of classical times and the Middle Ages down to the abstract speculations of our own day. None has survived the test of time, and we may be fairly sure that current theories will, in the near or distant future, suffer the same fate. But they nevertheless represent a particular milestone in the advance of human thought, and as such are worthy of consideration.

Bearing this reservation in mind, it can still be claimed that the number of gratuitously 'metaphysical' speculations has already been to some extent reduced. The determination of the age of the Earth, for example, was long ago transferred to the much more certain realm of experimental physics; that of the age of the Universe is beginning to be approached by different and not wholly speculative routes; and our knowledge of the evolution of the stars had made great strides since the subatomic sources of stellar radiation were properly understood. But pure speculation still dominates the realm of cosmogony properly so called — the origin and formation of the Universe and of its constituent stars — and no generally accepted account of this has yet been proposed. Even during the past 50 years an enormous mass of speculation has been published, and it would be a waste of time even to attempt a detailed survey of this; we will concentrate, rather, on a few of the most recent theories which are in not too marked disagreement with the results of observation — when, that is, such a juxtaposition of theory and observation is possible.

We shall see that, in their main outlines, the majority of these theories rely for support on three fairly recent discoveries:

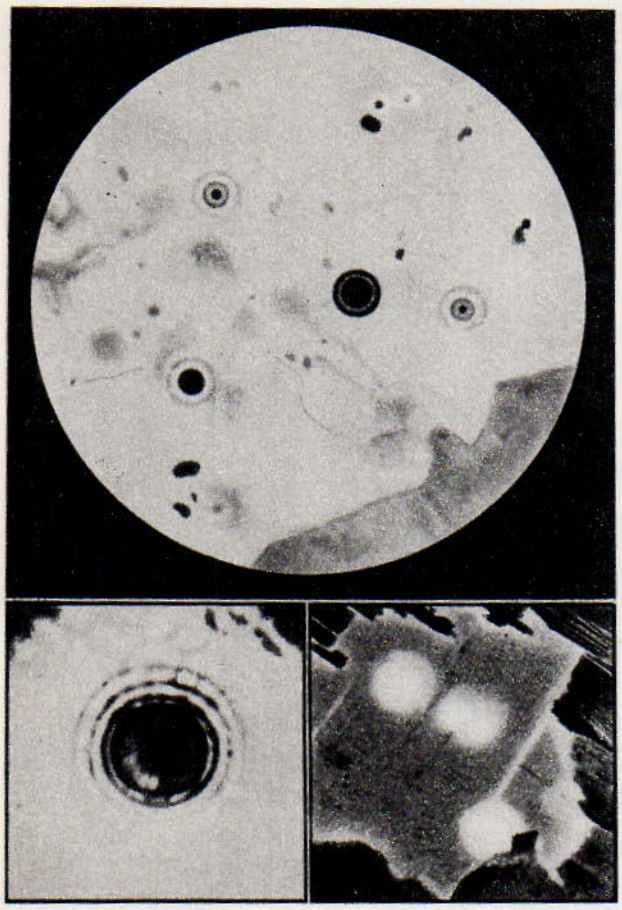


Fig. 717 — Pleochroic haloes, seen under a miscroscope, are the traces left in certain rocks by radioactive inclusions; their analysis permits the age of the rocks to be estimated.

1. The expansion of the Universe (provisionally accepted as real).
2. The universal preponderance of hydrogen, the simplest element, in both stars and interstellar matter, and the very considerable mass of this diffuse interstellar material.
3. The rôle played by the energy which is radiated by the stars, and its source in nuclear reactions.

These three groups of facts, only established within the last thirty years or so, lead to a conception of the past and future of the Universe which is very different from that held even a quarter of a century ago. But before embarking on an account of this, we must first see what is known concerning the age of the Earth and of the Universe.

The age of the Earth and of meteorites. — The age of the Earth has been fairly accurately determined by means of the radioactive minerals in its crust. We know that the majority of the heaviest elements are unstable and disintegrate spontaneously into lighter elements with the liberation of α-particles (helium nuclei), β-particles (electrons) or γ-rays. The lighter elements so formed are generally unstable also, and they in their turn disintegrate (after a shorter or longer interval); this process is repeated until finally a stable element is formed. In every case this stable end-product of the three radioactive families (starting from uranium, thorium or actinium) is lead.

All along the radioactive ladder from uranium or thorium to lead, helium (in the form of α-particles) is liberated. If the disintegration continues for a long enough period, 1 gm. of uranium will ultimately be transformed into 0.865 gm. of lead and 0.135 gm. of helium:

$$U_{92}^{238} \longrightarrow Pb_{82}^{206} + 5\ He_{2}^{4}$$

Now the speeds at which these elements disintegrate have been very precisely determined, and it has further been proved that the process can be neither accelerated nor decelerated by varying the physical conditions (temperature and pressure), or by the chemical reactions into which the radioactive elements of the Earth's crust can or do enter: such chemical raactions involve only the electron shells of the atoms, leaving their nuclei undisturbed. It has been established, for example, that each year 1.5×10^{-10} of *any* mass of uranium is converted into lead. Thus lead (and also, in theory, helium) will accumulate, slowly but inexorably, in every piece of uranium ore:

TIME (10^9 years)	PROPORTION (%) BY WEIGHT		
	URANIOUM	LEAD	HELIUM
0	100	0	0
1	86.5	12	1.5
2	75	22	3
3	65	30	5
5	50	43	7
∞	0	86.5	13.5

It will be appreciated that by means of this relation it should be possible to make some sort of estimate of the age of a radioactive sample, simply by determining the relative quantities of uranium, helium and lead that it contains. In practice it is the proportion of lead that is particularly useful, since the helium, being a light gas, is often able to escape from the ore; in practice, too, things are a little more complicated than this simplified account suggests, since different radioactive elements are often found together in natural rocks. But the principle is not affected, and the corresponding calculations can be made. The derived age of the oldest rocks is of the order of 1.5×10^9 to 2×10^8 years. This represents the minimum possible lapse of time since the Earth's crust solidified. To obtain the age of the Earth as an individual planet one must add a further period for cooling, which is estimated at several 10^8 years, giving the probable age of the Earth as something in the neighbourhood of 3×10^9 years.

Fairly recently (1946) this value has been improved by a remarkable method developed by the English geophysicist, A. Holmes. This also is based on the examination of the lead associated with radioactive decay. Without going into all the details, it involves the analysis (by means of a mass spectrograph) of the isotopes in samples of ore such as galena, obtained from geological strata whose age is already known. Naturally occurring lead is a mixture of several different isotopes, whose atomic weights are 204, 206 and 208; and the proportional abundance of these isotopes, besides varying with the nature of the sample, depends partly on the isotope constitution of the original lead in the Earth's crust and partly on the contributions resulting from radioactive decay; these modify the relative proportions of the different isotopes in a manner which can be calculated. In this way Holmes was able to deduce from the currently observed abundances of the different isotopes in rocks of various ages, both their initial proportions in the original lead and also the epoch at which these proportions began to be modified by contributions of radioactive origin. The values so obtained range from 2 to 4×10^9 years, but are

Fig. 718 — THIS PHOTOGRAPH OF THE GREAT NEBULA IN ANDROMEDA SHOWS US WHAT THE NEBULA LOOKED LIKE 1,500,000 YEARS AGO – THE TIME TAKEN FOR ITS LIGHT TO REACH US. The observation of distant nebulae carries us far into the past. *(Photo, C. D. Shane, Lick Observatory, 20-inch astrograph; 23 September 1946)*

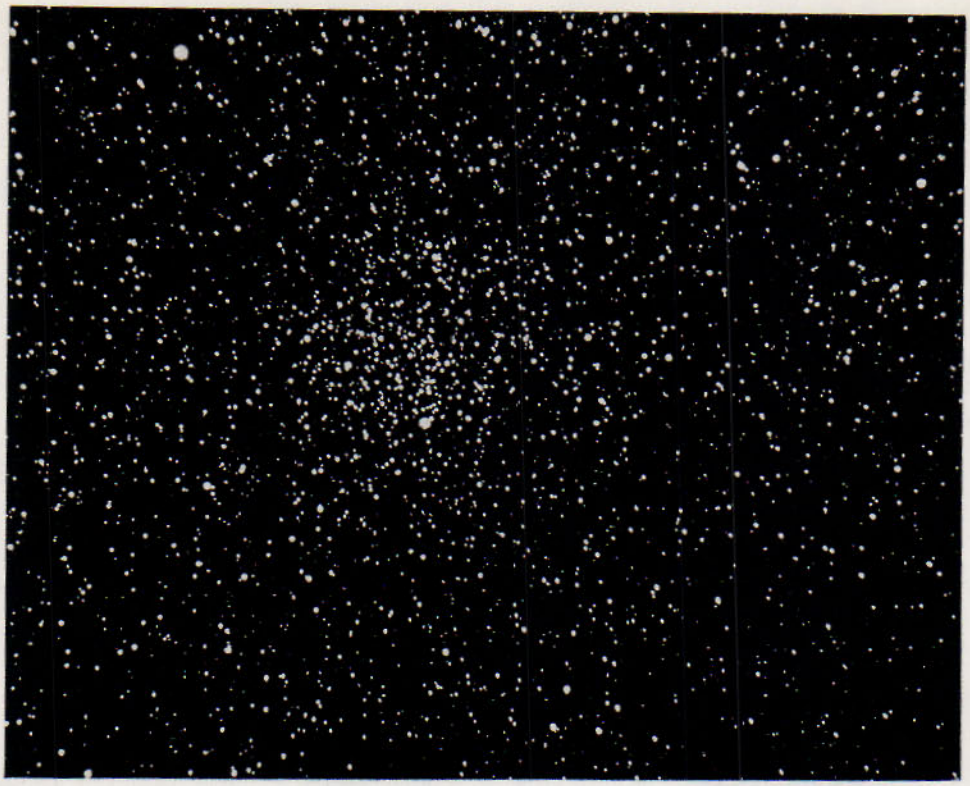


Fig. 719 — OPEN CLUSTER IN PUPPIS (M 46). The existence at the present time of a large number of condensed open clusters in the Galaxy is an indication of the comparative youth of the Universe. *(Photo, Jordan, Yerkes Observatory, 24-inch telescope)*

strongly concentrated around 3.3×10^9 years. Holmes finally concluded that the most probable value of the age of the Earth is 3.25×10^9 years.

It is interesting to note that the analysis of the helium, uranium and thorium content of meteorites, by Paneth and others, has yielded ages of the same general order. Though they are less precise, ranging from 10^8 to 3×10^9 years, this result may indicate that the Earth and meteorites have a similar origin — and certainly a comparable maximum age.

We shall see shortly that this similarity crops up in the most unexpected places, and that numerous independent lines of enquiry even agree in attributing an age equal to or slightly greater than the Earth's to the Universe as a whole. Only 25 years ago, this result would have been unthinkable.

The age of the Galaxy and of the Universe. — Before approaching this question it might be as well to consider for a moment what we mean by the 'age' of the Universe. What we do not mean is the date of a hypothetical Creation, on which the Universe was brought forth from nothing. On the contrary, in assigning a particular age to the Universe we are simply saying that for such-and-such a period it has existed in a form which does not differ too radically from that observed today, and which can logically be apprehended within the limited framework of our present-day knowledge. The fact that we cannot, today, probe further back in time does not, of course, mean that the Universe did not previously exist, but only that it is utterly impossible for us to envisage the nature of any hypothetical antecedent state. While this does not imply that there was no antecedent state, we are warned that any hypothesis regarding it must be treated as a metaphysical *jeu d'esprit*, devoid of scientific weight. For this reason, astronomers have tended to drop the expression 'the age of the Universe', with its over-ambitious metaphysical implications, in favour of 'the time-scale' of the Universe; this has the additional advantage of facilitating the juxtaposition of the 'long time-scale' (which was universally accepted 25 years ago) and the 'short time-scale', which is more widely accepted today.

At one time it was thought that since the age of the Earth is about 3×10^9 years, the stars and the Sun (which gave it birth) — and, *a fortiori*, the galaxies — must be at least several thousand times older than this. The ages attributed to the stars and galaxies were, in fact, roughly proportional to their sizes, and a number of miscellaneous observations were interpreted as indicating the validity of what is now known as the long time-scale.

But about 1930—32 the first doubts were cast upon the majestic lethargy of this evolution by the discovery of the expansion of the Universe. It will be remembered from the last chapter that the relation between the observed radial velocities and the distances of the galaxies implies that only 2×10^9 years ago the Universe must have been condensed into a very small volume — approximately, indeed, to a point. At that epoch the expansion is envisaged as having got under way, and it was very tempting to identify the beginning of the Universe with the beginning of its expansion. Today this line of reasoning seems less conclusive, particularly as the actual fact of expansion at the apparent velocity observed is still not firmly established. Furthermore, other indications of a shorter time-scale began to accumulate between 1934 and 1937, and today they constitute an impressive body of evidence. The most important of these indications of a 'short' time-scale (it would be premature to call them proofs) were in 1946 presented by the American astronomer, B. J. Bok, as follows:

1. The age of the Earth and of meteorites is, as we have just seen, about 3×10^9 years. It should also be noted that

Fig. 720 — BARNARD'S DARK NEBULA, B 92, IN SAGITTARIUS. Recent investigations have shown that the existence of dark nebulae in the Galaxy seems to favour the short time-scale. *(Photo, Mount Wilson Observatory, 100-inch telescope; 1921)*

— as has been pointed out by H. N. Russell — the mere existence of radioactive elements whose half-life* is of the order of 5×10^9 to 10^{10} years sets an upper limit of several 10^{10} years to the age of these substances.

2. The study of the dynamics of the more open galactic clusters (particularly by H. Mineur in France and Bok in America) has shown that those with a density of less than 0.1 Suns per cubic parsec (in the case of spherical clusters) are not stable, but are being rapidly dispersed by two agencies: the gravitational perturbation of nearby non-cluster stars, and the differential rotation of the Galaxy (see Chapter 16). The moving clusters in Taurus and Ursa Major are examples of such clusters in process of dissolution, and it can be calculated that the process will be completed within 19^9 years in the case of the Ursa Major cluster, and 3×10^9 years in that of the Hyades, which is the more condensed of the two (0.25 Suns per cubic parsec). Now the former has already practically ceased to exist as a cluster, while the latter is relatively little dispersed. It would therefore seem that 2×10^9 to 3×10^9 years is the significant period for the lifetimes of these galactic clusters.

3. The dynamical study of the more condensed galactic clusters leads to the same conclusion by a different route. Chandrasekhar has shown that in the case of clusters like the Pleiades and the Perseus double cluster, as well as the globular clusters, where the spatial density is much higher than the critical value, the gravitational influence of other galactic stars is negligible, but that the mutual interactions of the cluster stars themselves will lead to the progressive disintegration of the cluster at a rate which can be calculated. The cluster would thus slowly become more and more impoverished, and Chandrasekhar has calculated that the mean lifetime of a galactic cluster such as the Pleiades is as little as 3×10^9 years (compared with 10^{12} years for the globular clusters). The continued existence of large numbers of condensed open clusters is thus an indication of the relative youth of the Universe.

4. According to Chandrasekhar, the statistical investigation of wide binaries also provides evidence which is unfavourable to the long time-scale. He has shown that, under the influence of passing stars, wide pairs will part company altogether in a period of several 10^9 years. If the process of separation had been going on for a very long time, therefore, there should be many more very wide pairs than in fact there are. It appears, on the contrary, that this process of dissolution has hardly begun. On the other hand, we do not known that they were all formed simultaneously, and in any case the tally of very wide pairs is still far from complete.

5. Certain facts about the interstellar medium likewise appear to favour the short time-scale. In particular, it seems that the dark nebulae are unstable and must be dissipated (by diffusion and by the gravitational stresses of the rest of the Galaxy) within a time-span which varies from 10^8 to 10^9 years according to their mass. Now, as we have already seen, they are still to be found in great numbers, whence it would seem that they have not been in existence for a long enough time to disperse (unless some compensatory mechanism is simultaneously re-forming them).

6. The expansion of the Universe would also, in theory, constitute strong evidence for a short time-scale, since the onset of the expansion could not have occurred more than about 2×10^9 years ago. But, as already pointed out, the reality of this expansion is still in doubt.

7. Finally, the theoretical investigation of stellar evolution in the light of recent discoveries in atomic physics has shown that, on the long time-scale, we should be able to observe B, A and F type stars with absolute magnitudes of +2 or +3, and masses equal to that of the Sun. The absence of such stars from the Hertzsprung-Russell diagram appears to constitute a very strong argument in favour of the short time-scale.

When we come to consider the question of stellar evolution in greater detail we shall learn that it can be assimilated to the short time-scale without any difficulty; the white dwarfs alone seem to be too old to fit into this scheme, and their existence today has caused many a headache among the theoreticians. There are other phenomena, also, which are still regarded as being indicative of the long time-scale — for example, the distribution of extra-galactic nebulae within the great clusters, which Zwicky has shown to be analogous to that of the molecules in a gas sphere in isothermal equilibrium (see Chapter 17). To redress the balance, there are still others which suggest that the age of the Milky Way cannot exceed several 10^9 years. For example, the equipartition of kinetic energy, which is approximately realised in the Galaxy by stars of all spectroscopic types, was for a long time thought to favour a long time-scale; but since important exceptions have been discovered, it might equally well be adduced as evidence for the short time-scale, as had been argued by H. Mineur, when taking into account the diminution of stellar mass by radiation.

To sum up, many and varied considerations tend to suggest that the present state of the Universe cannot have continued for more than several 10^9 years. Thus the origin of the present phase of the Universe must be sought in events that — if the quite accurately determined age of the Earth is accepted as an indication of their temporal remoteness — occurred some 3.5×10^9 years ago.

We must now consider some of the recent theories as to the nature of these events, and the manner in which they might have precipitated the present-day Universe.

The origin of the Universe, and the formation of atoms. — The enormous progress that has been made in nuclear physics during the past 20 years, including our greatly increased understanding of nuclear reactions, has enabled us to tackle seriously the question of the origin of the chemical elements, and to suggest an explanation of why they occur in exactly the proportions they do, rather than in totally different ones. We know that hydrogen is not only the simplest of the elements but also much the most abundant, both in the stars and in interstellar space; helium, the next simplest, is also widely distributed throughout the Universe; then come various light elements, whose proportional abundances are already very much smaller — carbon, nitrogen, oxygen, sodium, aluminium, etc.; elements of moderate atomic weight (calcium, titanium, iron, nickel, etc.) are still less abundant; and finally, the heavy elements (such as silver, gold, lead and the naturally radioactive elements up to uranium, the heaviest) are very rare, and occur in progressively smaller quantities. At the same time, it is well known that some isotopes are very abundant while others are equally rare. The oxygen of our atmosphere, for instance, consists of three different isotopes, whose atomic weights are 16, 17 and 18, and whose relative abundances are respectively 99.76%, 0.04% and 0.20%. Similar differences of isotope abundance are encountered in other elements.

What is the reason for this progressive — and, if we ignore certain secondary exceptions, quite general — rarity of elements with increasing atomic weight? Why are some isotopes so much more abundant than others? How, and under what conditions, were they formed? Tentative answers to these and similar questions have in recent years been offered by various theoretical astro-

* I. e., the time required for one atom in every two to disintegrate.

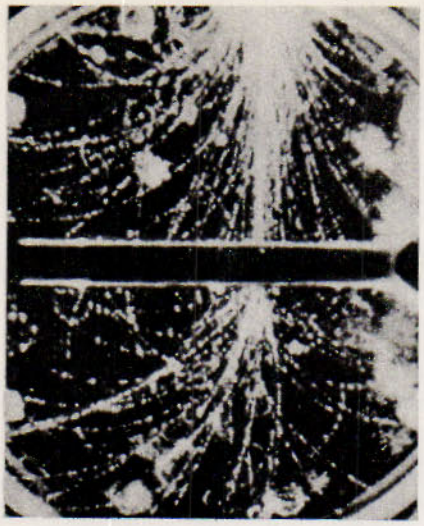


Fig. 721 — A FINE SHOWER OF PARTICLES LET LOOSE IN A WILSON CLOUD CHAMBER BY COSMIC RADIATION. *(Photo, Neddermeyer and Anderson; 1934)*

physicists, notably C. von Weizsäcker in Germany and Chandrasekhar and Henrich in America. Two reasonably well established facts are relevant to the problem: in the first place, it is nowadays regarded as highly likely that the various chemical elements were originally built up by the combination of the simplest fundamental particles (neutrons, protons and electrons); and secondly, in the light of our knowledge of stellar interiors it seems unlikely that the formation of heavy elements is possible in the present-day Universe. However, there is not general agreement on this latter point: Hoyle, for example, in 1946 propounded the view that the heavy elements could still be produced in the cores of white dwarfs. But we shall not stop to discuss this hypothesis, confining ourselves to those that have been more widely considered during recent years.

The formation of the elements must have antedated that of the stars; it must have occurred during a pre-stellar stage of the Universe, characterised by the exceedingly high temperatures and pressures which are necessary to produce the requisite nuclear reactions. Some confirmation of the hypothesis of a hyperdense pre-stellar state of the Universe is provided by its apparent expansion, which implies that several 10^9 years ago its entire material content was condensed into an extremely small volume.

Calculation (Chandrasekhar, 1941, in particular) of the nuclear reactions and equilibria that are possible between neutrons, electrons and α-particles under these conditions has in fact accounted fairly accurately (more could not be expected of first attempts) for the relative abundances of the lighter elements and their isotopes up to an atomic weight of about 40, assuming a pre-stellar environment whose temperature was of the order of 5×10^9 to 10^{10} degrees, and density about 10^7 times that of water. But even these extreme conditions could not permit the formation of the heavy and very heavy elements, and Chandrasekhar was forced to invoke an earlier and even denser state, in which the temperature exceeded several 10^{10} degrees; this material, beginning a progressive expansion, would contain heavy atoms that then encountered lower temperatures, and in the same way, at still lower temperatures and pressures, the lightest elements would in turn be established. Finally, as the expansion of the pre-stellar Universe continued, the proportional abundances found today would be established by the series of nuclear reactions which are still going on in stellar interiors. The degree of condensation of the Universe in its pre-stellar stage exceeds the bounds of imagination: it has been calculated that if its density were raised to 10^7 times that of water, the entire mass of the Universe could be contained in a sphere whose radius is less than 0.01 parsecs, or less than 50 times the present size of the Solar System.

These views have nevertheless run into difficulties, and in 1947 van Albada in Holland and P. Lacroute in France showed that the simultaneous formation of all the elements, including the heaviest, could be accounted for more simply (and without having recourse to such fantastic temperatures) by hypothesising a pre-stellar state in which the Universe consisted of a large number of aggregations comparable to the present-day white dwarfs, at whose centres the heavy elements were formed, to be subsequently dispersed.

However that may be, being able to identify the main outlines of the process whereby the different atoms were built up from the fundamental particles of Nature, and also to explain their relative abundances, is a remarkable enough achievement, and justifies a certain degree of confidence in the theories that envisage a pre-stellar state of the Universe. But cannot we go even further, and envisage a still earlier primal state in which the density of matter is pushed to the furthest limit that is conceivable? It is at any rate theoretically possible to consider a state of condensation such that the atomic nuclei are actually in contact with one another, fusing into a single gigantic nucleus — the 'primal atom'. We have already seen that high enough pressures can destroy the individualities of the atoms (Chapter 15), reducing them to the condition of degenerate matter; so, too, the super-pressures here envisaged would destroy the individuality of the atomic nuclei themselves, and reduce the matter to a 'nuclear state' whose density is the absolute maximum that can be conceived. This density is of the order of 10^{12} times that of water.

Thus, a few 10^9 years ago, the entire material Universe (whose mass may be put at about 10^{21} times that of the Sun) may have been concentrated into a unique and gigantic super-nucleus, with a density 100,000 times greater even than that of the pre-stellar state described above, and a diameter not exceeding that of the Solar System today — less than 10^{10} miles.

Thus by a somewhat different route we have arrived back at the Abbé Lemaître's conception of a primal atom, the essential points of which may be summarised as follows. It is supposed that the primal atom — the original nucleus — was endowed with a super-radioactivity which engendered, by successive disintegrations, the elements existing today, including the radioactive elements; these are the still active descendants of heavier and less stable parents which have long ago vanished from the scene. The cosmic rays, about which so much has been written, are simply the super-γ-radiation that was emitted during this phase of super-radioactivity: generated at the dawn of Time, they have ever since continued to circulate throughout the spherical Universe of relativistic theory, as though in a closed vessel. Thus is explained both their prodigious energy (more than 10^9 electron-volts) and their isotopy — the fact that they appear to arrive indiscriminately and equally from every direction.

Lemaître's additional assumption of primal super-radioactivity no longer appears to be an indespensable feature of the hypothesis, for, as we have seen, a beginning has already been made to explain the existence of the different elements without it. A simple assumption regarding the nature of the primal nucleus would be all that

is required, and in this connexion it was suggested some years ago by the French physicist, F. Joliot, that the simplest of all is that the nucleus consisted solely of neutrons which, it seems, could give birth to the other fundamental particles by spontaneous transmutation.

Whatever we may think of these ultimate hypotheses, there is no difficulty in accepting the fact of the inherently explosive nature of the primal atom and that from that first moment (beyond which our ignorance is complete) there must have set in the first violent and rapid phase of expansion during which, while the Universe was still in the pre-stellar stage, the various elements existing today were formed.

We shall see how, from this starting point, the Abbé Lemaître has been able to explain the main outlines of the evolution of the present-day Universe.

Formation of stars and galaxies. — In order to understand the evolution of the Universe from the primal nucleus to its present state, certain principles of the theory of the expanding Universe must be recalled.

According to this theory, the equilibrium of the Universe and the direction of its evolution (expansion or contraction) are determined by two fundamental forces — one of attraction, which is inversely proportional to the square of the distance (universal gravitation), and one which is proportional to distance and may be either an attraction or a repulsion, according to the sign and magnitude of the cosmological constant in Einstein's equations (see Chapter 17). The observed expansion of the Universe shows that the sign of this constant is positive and that the second of these fundamental forces is therefore one of repulsion. Given exactly the right density of matter, the two antagonistic forces will balance one another, and we shall have a static universe in unstable equilibrium — the Einstein universe. There exists a simple relation between this critical density and the cosmological constant. If the mean density of matter throughout the Universe is less than the critical value, cosmic repulsion will overcome gravitational attraction, and the Universe will expand — as appears to be happening to the Universe at the present time. If, on the other hand, the mean density exceeds the critical value, contraction will set in and will continue until a state of equilibrium is achieved; this will be determined by the forces due to random motions of the material content, and to radiation, and is realised at the present time in intragalactic space, i.e. in the interiors of the galaxies.

It occurred to the Abbé Lemaître that regions of intermediate density might still exist in the Universe where exact equilibrium between gravitation and cosmic repulsion is realised; the density of such a region would give us the value of the cosmological constant, which is not specified by theory. He believed that he had located such regions in the irregular clusters of nebulae (such as the Virgo cluster, for example) whose central condensation is not so marked as to suggest the action of a dominant force of attraction.

Since the populations and diameters of these clusters are known, it can be established that the mean separation of the galaxies within them is about 200,000 lightyears; this is 10 times greater than in extragalactic space generally, where the average distance from galaxy to galaxy is of the order of 2×10^6 lightyears. From the (approximately) known masses of the galaxies it can be calculated that the corresponding mean densities of matter are respectively, and very roughly, 10^{-27} and 10^{-30} gm./cm.^{3}, or 1,000 atoms per cubic metre inside a cluster and only 1 atom per cubic metre throughout the Universe as a whole. These infinitesimally small values will be slightly augmented if account is taken of the fact that the mean density of the interstellar material within the Galaxy is estimated as 10^{24} gm./cm.^{3}, corresponidng to 10^6 atoms per cubic metre.

Taking all these facts into account, the Abbé Lemaître pictures the past history of the Universe somewhat as follows. In the initial explosive phase the primal nucleus expanded violently, no doubt breaking up to some extent, and giving birth to the different kinds of atom as explained above. But owing to the enormous density of the system its velocity of expansion would be braked progressively by the vastly more powerful force of gravitation. The explosive pre-stellar stage would therefore be followed by one of deceleration, during which the mean density of matter would approximate to and finally attain the critical value required for equilibrium. In this state — which approximates more or less closely to the Einstein universe — variations of density due to local random motions would give birth to clouds of matter whose density was greater than the average — the galaxies. An analogous process on a smaller scale would have given birth to the stars. Since the critical density would be exceeded inside these condensations the force of gravitation would overcome that of repulsion, and from this point of view they would be stable. It was thus during its decelerating phase — perhaps 3×10^9 or 4×10^{10} years ago — that the present stellar organisation of the Universe was established. Thereafter, with expansion continuing and the mean density falling below the critical value, cosmic repulsion would triumph over gravitation, and the 'expansion of the Universe' would be on its way. The galaxies that had condensed during the preceding phase would separate progressively from one another, and this would in turn increase the preponderance of cosmic repulsion over gravitational attraction. The only exceptions to this would be certain atypical regions (caused by rare but natural fluctuations during the Einstein phase) in which the unstable equilibrium would be prolonged and would eventually survive almost to the present time within the great clusters of extragalatic nebulae.

The difference between the mean separations of galaxies in clusters and in the general field shows that since its passage through the equilibrium state the Universe has roughly doubled its size. Its mean density at the present time is approximately 1,000 times lower than it was then, and the repulsive force is so much stronger than that of attraction that the latter may now be disregarded. In other words, expansion has already carried the Universe virtually to the condition of the empty de Sitter model (cf. Chapter 17). But in de Sitter's universe the relation between radius and cosmological constant is known, and the latter can therefore be derived from the observed velocity of expansion.

Now these two independently derived values of the cosmological constant (from the equilibrium density of the clusters of galaxies, and from the velocity of expansion of the Universe) are in excellent agreement with one another. This agreement clearly adds considerable weight to the theory which has just been described and which may then have appeared to be entirely arbitrary.

It therefore follows from the theory of expansion that at the epoch of equilibrium density the radius of the Universe was about 10^9 lightyears, and that it is now of the order of 10^{10} lightyears. This account of the general evolution of the Universe can be reduced to the data given in the following table, and represented diagrammatically as in Fig. 722.

It remains to say something about the conditions under which the stars and galaxies were formed from the chaos of the deceleration phase of the Universe, 3 to 4 thousand million years ago. The question of priority immediately presents itself: which came first, the galaxies or the stars?

EVOLUTIONARY STAGE	EPOCH (—10^9 years)	RADIUS OF UNIVERSE (lightyears)	DENSITY (water = 1)
Primal nucleus	?	10^{-4}	10^{12}
Formation of atoms	?5	10^{-2}	10^{7}
Formation of stars	?4 to ?3	10^{9}	10^{-27}
Today	0	10^{10}	10^{-30}

or were they formed simultaneously? Various answers have been given to these questions. Gamow and Teller, for example, believe that the formation of stars preceded the separation of the galaxies from the undifferentiated primal mass; others hold that the galaxies separated first, the subsequent formation of stars being confined to their gaseous interiors. A mechanism whereby the birth of stars might occur under these circumstances has recently been proposed by F. Hoyle, and depends on the differential rotations of the galaxies. This rotation could have been set up during the epoch of their initial condensation, as a result of direct but asymmetrical collisions, or more probably of grazing contacts between neighbouring embryonic galaxies during the equilibrium phase, when they were very close to one another. A similar process will be described later in connexion with the birth of planets.

The latter alternative seems to receive some support from the discovery of the galactic globules (see Chapter 16), which are thought to be stars in process of formation. It is nevertheless by no means out of the question that some types of star are older than others — some having been formed in all parts of the Universe before or during the splitting up of the galaxies, others within the galaxies themselves after their separation. Indeed we have already seen that for some time now it has been recognised that there are two categories of stellar population: Baade's Population I, which is found only in conjunction with Population II and even then is confined to the arms of the spirals (characterised by blue supergiants), and Population II, which is widely distributed throughout all the galaxies, and is alone represented in the globular clusters and elliptical nebulae.

We shall see later that the supergiants characteristic of Population I would seem to have very short lives compared with other types of star, and their existence at the present time therefore suggests that their formation is still going on in the heart of the interstellar medium. The lifetimes of the dwarfs, on the other hand, and even of the giants, are considerably greater than the age of the galaxies. It therefore appears very probable that they were formed at an earlier date than the galaxies, and that they can still look forward to a long future.

Before turning our attention to this question of the future prospects of the Universe, we must first review some of the theories that have been proposed to account for the formation of planetary systems, and in particular that of the Earth, whose age indicates that it must have been born during this same critical period in the evolution of the Universe.

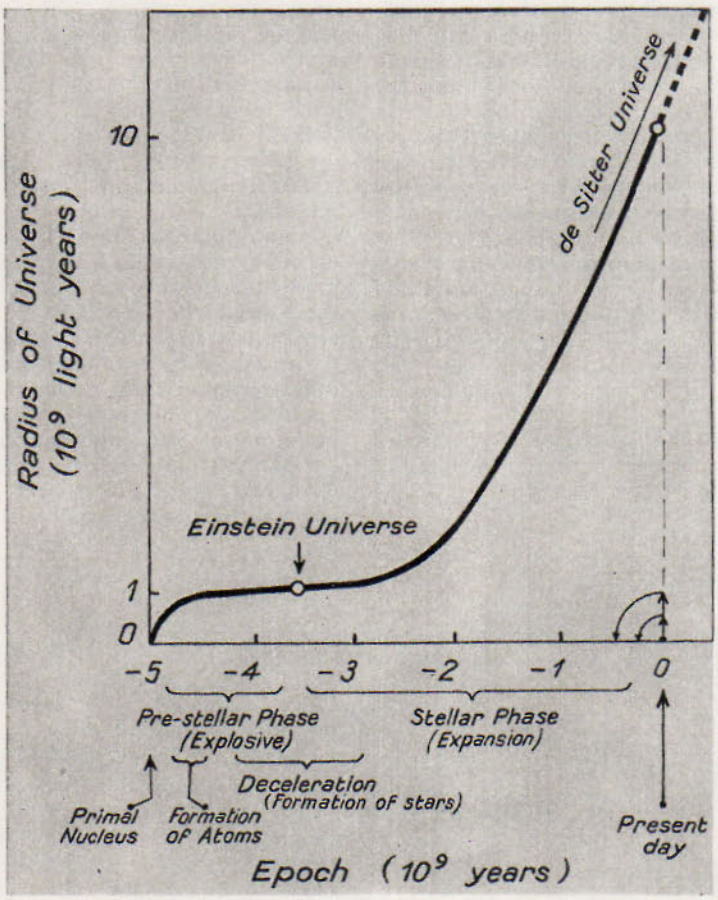


Fig. 722 — DIAGRAM SHOWING THE PRINCIPAL STAGES IN THE EVOLUTION OF THE UNIVERSE, AS CURRENTLY ENVISAGED. The two arrows and their curved extensions (bottom right) represent, on this scale, the space — and time — penetration of the Mount Wilson 100-inch telescope and the Mount Palomar 200-inch.

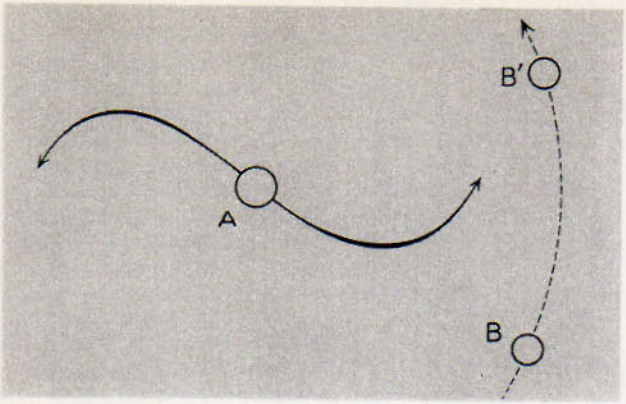


Fig. 723 — THE TIDAL THEORY IN COSMOGONY. Diagram showing the changing direction of the attraction exerted by the passing star B on the star A, from which it draws out a filament of matter.

The formation of planets. — Although the origin of the Earth and other planets has long been the subject of speculation — until recent times, indeed, it constituted the whole of cosmogony — it is still far from being satisfactorily explained. In view of the proximity of the planets and the comparative ease with which their origin can be imagined, it may seem paradoxical to attack the problem of the origin of the stars and of the Universe as a whole when that of our nearest neighbours, and of the Earth itself, is still uncertain. There is, however, a fundamental reason for this: we can observe millions of stars and galaxies, and can therefore compare and classify them, distinguishing general laws and sorting out the essential facts from those that are only secondary or accidental; but we are, in practice, acquainted with only a single planetary system, and therefore find it much more difficult to distinguish which of its many characteristics are significant. Those which are generally accepted as being essential to the problem of origin are:

1. The usually small inclinations of the orbital planes to the mean plane of the system (asteroids and Pluto excepted); the latter is itself not quite coincident with the equatorial plane of the Sun.

2. The fact that planets and satellites both rotate and revolve in the same direction (Uranus and certain satellites excepted).

3. The near-circularity of the planetary orbits (Mercury and Pluto excepted) and their distribution in the ordered manner described by Bode's law (Neptune and Pluto excepted).

4. The concentration of practically the whole of the system's rotational momentum in the giant planets.

Among the many cosmogonies that have been proposed in the past, one of the most important historically was that of Kant and Laplace, further developed by Roche, Faye and others in the nineteenth century; this supposes that the Sun was once diffuse, nebulous, and very much larger than at present, rotating rapidly and at the same time condensing; as it contracted it is supposed to have ejected successive rings of diffuse matter which later condensed into planets. This theory is incapable of accounting for such facts as the rotations of the planets, the inclinations of their orbits to the solar equator, and retrograde motions within the system; it is furthermore based on assumptions which are now recognised to conflict with certain established laws of physics. Since the turn of the century it has been generally abandoned.

Various theories have been based on the assumption of a more or less tangential collision between the Sun and another star. This was first considered by Buffon in the eighteenth century, and in our times has been developed by H. Jeffreys and Lyttleton in England. But these hypotheses have likewise failed to stand up to the criticism levelled at them, and the extraordinarily low density of stars in space makes the occurrence of direct collision highly improbable.

On the other hand a group of theories involving, not an actual collision, but only the near approach of another star and the Sun — developed, in particular, by Moulton, and later by Jeans in the early years of this century — merits closer scrutiny. Such a near appraoch of two stars may be regarded as a semi-collision, since their mutual gravitational attractions would produce a marked deformation of their paths into two hyperbolic arcs, and the effect might be practically indistinguishable from that of a direct elastic impact of the two bodies, from which they would rebound in opposite directions.

But the close approach of two stars will have another important consequence: the enormous tidal forces set up in them by their mutual gravitational attractions at the epoch of minimum separation would result in jets of matter being drawn out from the crests of the tidal bulges; these would be stretched out into the form of curved arms whose shape would be determined by the changing directions of the two perturbing bodies (Fig. 723). As they separated from each other the tidal bulges would gradually collapse and the base of the jets would fall back into the Sun. But their outer parts, having acquired a sufficient rotational velocity, would remain separated from both stars, and would break up into individual 'drops'; these would condense rapidly, while continuing

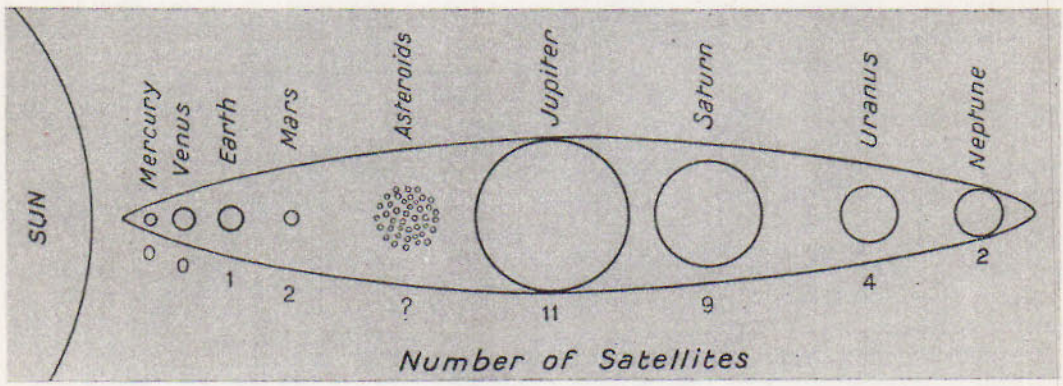


Fig. 724 — DIAGRAM SHOWING THE DISTRIBUTION OF THE PLANETS WITHIN A SPINDLE-SHAPED FILAMENT OF MATTER DRAWN OUT FROM THE SUN (JEANS'S COSMOGONY). Because of the uncertainty about its diameter, Pluto is not shown.

to revolve about the parent star, thus giving birth to planets.

This hypothesis has enjoyed a wide popularity during the last forty years. It accounts satisfactorily for the occurrence of the giant planets in the centre of the planetary region (Fig. 724), for the non-coincidence of the mean plane of the orbits with that of the Sun's equator (there being no reason why the path of the passing star should lie in the latter plane), and for the speed of the planetary rotations. Owing to the low probability of even this type of stellar encounter in the present state of the Universe, however, the hypothesis implies that planetary systems must be very rare, and confers on the Solar System a status which, if not unique, must still be most exceptional. But on the short time-scale this objection vanishes, especially if it assumed that the majority of the stars were formed at a very early stage in the development of the Universe; for they would then have been very much more closely packed than they are today, and near encounters would have been correspondingly more frequent. The conclusion that planetary systems similar to ours may be much more numerous than was thought 30 years ago may perhaps find observational support in the recent discoveries of 'planetary' companions to certain of the nearer stars (Chapter 13).

The theory nevertheless does not give a complete account of the Solar System as it is today. For one thing it can offer no explanation of the orbital relationship expressed in Bode's law. In this connection the theory formulated by the German astrophysicist, C. von Weizsäcker, in 1943 (which may be regarded as a logical development of the near-approach theory) is much more satisfactory.

Weizsäcker supposes that the planets were formed within a flattened gaseous envelope rotating about the Sun; this was either the remains of a cloud of cosmic material from which the Sun had condensed at an earlier epoch, or else consisted of matter wrenched from the Sun by tidal action during the near approach of another star, as envisaged by the Moulton-Jeans theories. The disc, whose mass was probably not more than one-tenth of the Sun's, would be unstable, tending to dissipate rapidly into space — possibly within a period not exceeding several 10^7 years. The dissipation would be particularly rapid in the case of the lightest gases, hydrogen and helium, which explains their relative rarity in the planets, compared with the Sun. Moreover the innermost regions of the cloud would tend to fall back into the Sun, so that (hydrogen and helium constituting, as we have seen, at least 90% of the Sun's mass) only about 10% of the total mass of ejected matter would finally remain in the gaseous envelope as the raw material for the formation of planets. This agrees well with the observed combined mass of the planets, which is roughly 10^{-3} that of the Sun.

Despite the general instability of the disc, internal motions would be set up which would be relatively stable and permanent: Weizsäcker was able to prove that the atoms and other particles of the disc, revolving about the Sun in circular or elliptical orbits, would very rapidly become organised (probably in a matter of centuries) into a number of individual cells (Fig. 725); each of these would rotate, whirlpool-fashion, about its own centre, and all these centres would themselves revolve around the Sun in circular orbits. Since adjacent cells could not encroach on one another the extent of each would be limited by the presence of its neighbours — laterally by those revolving in the same orbit, and radially by those that had developed in adjacent orbits, external and internal. In this way the disc would break up into a pattern of individual cellular whirlpools strung out in concentric

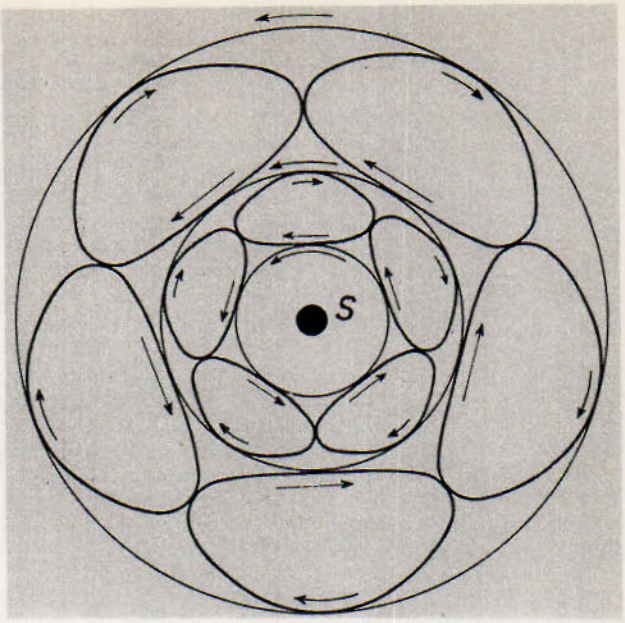


Fig. 725 — WEIZSÄCKER'S THEORY OF THE FORMATION OF THE SOLAR SYSTEM FROM CELLULAR VORTICES

rings about the Sun. Now, if it is assumed that each successive ring contains five such cells, it is found that the radii of the circles separating them would be proportional to the successive powers of a number very near to 2 (actually, 1.894), which conforms closely with the series known as Bode's law; in fact, the calculated figures for the planets from Mars to Uranus are in even closer agreement with observation than those given by Bode's empirical law. In this way was explained the approximate regularity of the distribution of the planetary orbits, which had for so long been a stumbling-block for cosmogonists. The agreement for the innermost and outermost planets — which would develop from the least stable regions of the gaseous disc — is less satisfactory, and a somewhat larger number of cells in each ring (7 to 10) has to be assumed.

It is a consequence of this hypothesis that the planets would be formed at the inner and outer edges of the concentric zones, where the whirlpool motions of adjacent cells would be opposed to one another; this opposition would give rise to local eddies whose direction of rotation

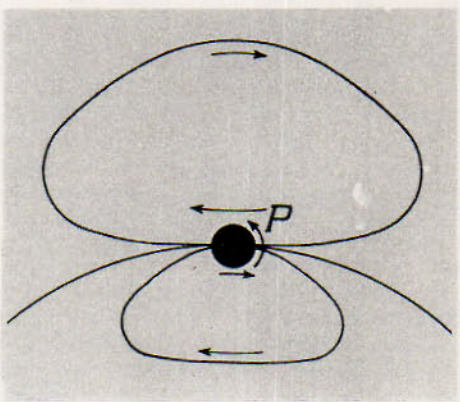


Fig. 726 — FORMATION OF A PLANET AS A RESULT OF THE INTERACTION OF TWO CELLULAR VORTICES, according to Weizsäcker.

would necessarily be the same as their direction of revolution (see Fig. 726); thus is explained the observed fact of the planets' directional unanimity of rotation and revolution. These smaller eddies would in their turn give birth to satellites by a like process, while the planets would condense by coalescence and the progressive capture of small particles by large; all this is described in detail by Weizsäcker, but cannot be further gone into here. It will merely be noted that the period of planet-formation must have been relatively brief (probably less than 10^8 years), because at the same time the cloud itself would be dissipating rapidly, as explained above.

Hence we can see that although there is, at the present time, no definitive and firmly established explanation of the birth of the planets, it is possible to give some account of its main outlines which agrees reasonably well with the facts of observation. But we still have to explain how the Earth, once its status as an independent entity was established, developed the form and structure that it shows today. To do this, we must turn our attention to the remote period of its history that extends from its birth down to pre-Cambrian times — the period of more than 2×10^9 years that preceded the considerably shorter one (10^9 years) which is the domain of the geologist and which culminates in the present day.

The Earth's prehistory. — According to the French geophysicist, A. Dauvillier, the history of that remote and lengthy phase of the Earth's evolution is as follows. when once the Earth had established itself as a separate entity, in the form of a gaseous sphere — which Dauvillier, accepting the tidal theory of Jeans, supposes to have been at a very high temperature — the elements began to separate out according to their atomic weights, the heaviest sinking to the centre. At the same time, chemical reactions began to come into the picture, with the consequent formation of molecules.

At a temperature of less than 3,000° the terrestrial core condensed very rapidly, first becoming liquid and then solidifying. At the same time the light metals (silicon magnesium, aluminium, calcium) entered into combination with the non-metalic elements (such as carbon, nitrogen, phosphorus and sulphur) to produce carbides, nitrides, phosphides and sulphides.

A little later, and at a somewhat lower temperature, oxygen became able to form stable oxides of all sorts (silica, magnesium oxide, alumina, etc.) which would be liquid at the surface of the core; when the temperature dropped to 1,700°—1,500° C, thick layers of silicates formed round the core, constituting what geophysicists term the 'sima' (*si*licates of *ma*gnesium) and the 'sial' (*si*licates of *al*uminium). The former, with densities ranging from 3.0 to 3.4, is about 700 miles thick, and the latter (densities 2.8—3.0) not more than 125 miles; basalt (density 3.0) formed a transitional layer between them. The Earth's crust, properly so called, comprises the most superficial and lightest layers of the sial; its composition is gneiss and granite, with a mean density of 2.8, and its total thickness is only about 40 miles. This granitic crust solidified first, forming a kind of gargantuan skin over the molten interior; this skin was immediately split and rent by its own contraction, breaking up into fragments that might be likened to enormous icebergs of granite floating in an ocean of liquid basalt; with a density of 2.8 (as against the basalt's 3), about 1/15 of these floating granitic islands would show above the surface — i. e. 2 to 3 miles, which is the present-day difference between the mean height above sea level of the continents and depth below sea level of the ocean beds. Dauvillier believes that the ocean deeps can be explained in terms of the local subsidence of the granitic skin, which would be followed by outpourings of basaltic lava from the interior. At the same time, the levels of vast areas were rising, under the action of vapours at high pressure escaping from the internal semi-liquid mass.

Thus the incandescent terrestrial globe (at a temperature of about 1,500° C) passed through a stage similar to that in which the Moon solidified.

The condensation of the crust, first to a liquid and then to a solid, was also responsible for the development of the particular atmosphere which the Earth has. The original atmosphere of hydrogen and helium would have been short-lived, since the Earth's mass is too small to have prevented the greater part of these very light gases from escaping into space; a small proportion of the original hydrogen was alone retained, owing to its chemical properties which allowed its fixation by the reduction of metallic oxides and by combination with oxygen to form water vapour. The pressure of this water vapour must have exceeded 300 kg./cm.² (roughly 400 lb/sq. in., or 300 atmospheres), for all the water that fills today's oceans was then contained in the atmosphere. Reaction between the water vapour and the previously formed carbides, nitrides, phosphides and sulphides produced an atmosphere of hydrocarbons (acetylene, methane, etc). and of gases like ammonia and phosphoretted hydrogen, somewhat reminiscent of the present-day atmospheres of the giant planets. The hydrocarbons in turn reacted with the water vapour, giving first carbon monoxide and then, at a lower temperature, carbon dioxide. Still later, when the temperature dropped below 100°C, the water vapour condensed and the Earth was subjected to a scalding deluge of unimaginable violence, which levelled the original contours of the crust and poured into the immense deeps which, as already explained, were ready waiting to receive it. Thus were formed the oceans, which, with subsequent minor modifications due to small-scale movements of the Earth's crust, have survived to the present day.

At an earlier epoch than this, when the temperature fell below the 500° mark (the magnetisation temperature of ferrous minerals), the jets of electrons that curve in from the Sun were responsible for the magnetisation of the ferrous deposits contained in the sial. The Earth's magnetic field, thus set up, thereafter protected the surface of the globe from further electronic impacts.

The cooling of the Earth's crust must have been completed more than about 2×10^9 years ago, by which time the Earth had already — astronomically speaking — assumed all its essential present-day characteristics. At that epoch the temperature fell below the boiling point of water, and the oceans were permanently established. But the atmosphere, which was then composed of carbon dioxide, nitrogen and water vapour, was still transparent to the Sun's ultra-violet radiation, and under the influence of the latter the carbon dioxide was able to combine with water to produce formaldehyde, whose polymerisation led to the formation of complex carbohydrates (glucose, starch, etc.) in the surface layers of the oceans. This process was soon brought to an end, however, for the oxygen liberated by the reaction was itself polymerised by the solar ultra-violet radiation, with the production of increasing quantities of ozone; this gas is opaque to the active wavelengths and soon prevented them from ever again penetrating to the Earth's surface.

Assuming all this, and taking into account the effect of electrical discharges in the atmosphere. Dauvillier showed that enormous quantities of gelatinous and colloidal organic matter could be formed in the surface layers of the oceans; he and the biologist, Desguin, believe that the subsequent appearance of life in this organic environment can be accounted for if appeal is made to solar ultra-violet

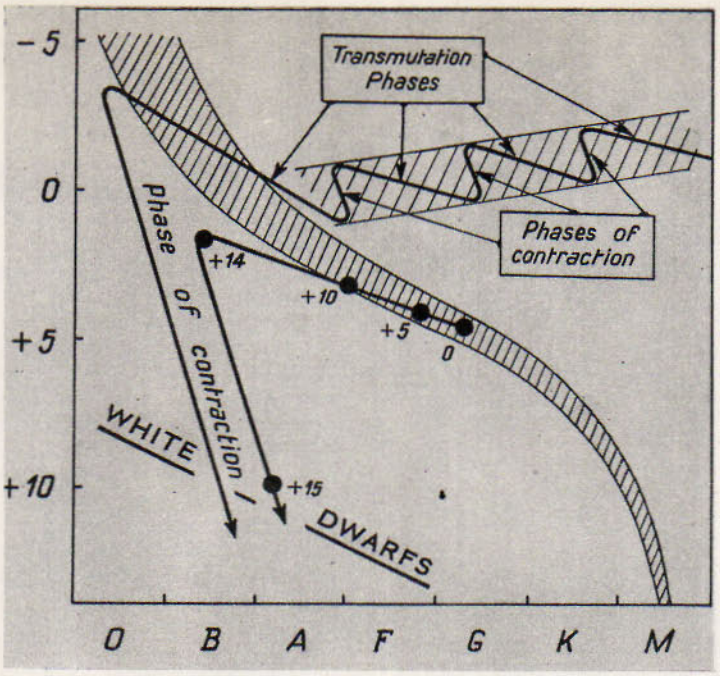


Fig. 727 — STELLAR EVOLUTION. Diagram showing the probable routes of two stars across the Hertzsprung-Russell diagram during the course of their evolution. That of a dwarf, the Sun, is marked off in periods of 5×10^9 years. The other star, a giant, undergoes alternate phases of contraction and transmutation.

radiation leaking through the atmosphere.*

But at this point we come to the beginning of the geological era, whose history lies outside the scope of this book.

We must now review the entire history of the stars and galaxies — past, present and, finally, future — in so far as its general outlines can be discerned, in an attempt to discover an explanation of their present properties and behaviour. At the same time we shall assemble those elements of the picture that will eventually enable us to glimpse something of the future of the Universe.

Evolution of the stars. — We have already indicated the main outlines of the development of the stars during the decelerating phase of the expansion of the Universe. But we still have to explain their present diversity and individual characteristics. The replacement of the long time-scale by the short has imposed the necessity of a complete revision of our ideas in this field.

In terms of the long time-scale, one could assume that the stars, having been formed by condensation in a diffuse gaseous nebula, started life as extremely low density red giants or supergiants; contraction would cause their temperatures to rise slowly, by Helmholtz's mechanism (Chapter 15), with the result that they would pass along the giant branch of the Hertzsprung-Russell diagram. At a certain point their densities would reach the value beyond which further contraction could not compensate for their radiation losses, and a much longer phase of cooling would then set in, during which they would retrace their steps through the spectroscopic types from B to M, shrinking steadily all the time, and moving down the dwarf branch of the diagram. After many 10^{12} years, old age would overtake them and, as advanced red dwarfs, they would have nothing further to look forward to save a final and irrevocable relapse into black dwarfs.

This picture of stellar evolution, which in any case provided no explanation of the existence of the white dwarfs, has been jettisoned in its entirety since the discovery of the nuclear reactions that provided the stars with their energy (see Chapter 15).

In this reassessment, the carbon-nitrogen cycle of Bethe and Weizsäcker plays a decisive rôle. It will be remembered from Chapter 15 that this cycle of reactions — which is kept going by carbon and nitrogen, and in the course of which hydrogen is converted into helium with the release of energy — is the mechanism whereby the majority of Main Sequence stars derive the energy which they radiate. G. Gamow has shown that the carbon-nitrogen cycle has a remarkable and surprising property — as the reserves of hydrogen in a star diminish, so its production of energy increases. This is due to the fact that the energy produced in the centre of a star is not transmitted in its entirety to the surface, whence it can be radiated into space; on the contrary, it is largely retained in the interior, whose opacity regulates the star's expenditure of energy (see Chapter 15). Now it has been discovered that helium, which is the end-product of the cycle, is more opaque to radiation than the hydrogen from which it is derived; accordingly, as the body of a star becomes richer in helium and poorer in hydrogen it will also become increasingly opaque; the store of radiation trapped in its central region, where the nuclear reactions take place, will therefore also increase. Hence the central temperature will steadily rise, thus increasing the thermal agitation of the nuclei; this in turn will increase their capacity for provoking nuclear reactions, and the

* Protein — the substance of living matter — is built up of simpler, but still very complex, units known as aminoacids; these in turn consist of carbon, oxygen, hydrogen and nitrogen atoms. In an attempt to reproduce the chemistry of the Earth's early environment (water H_2O, ammonia NH_3, methane CH_4 and carbon dioxide CO_2), the American, Millen, enclosed nitrogen, hydrogen, carbon and oxygen in a tube and exposed it to ultra-violet radiation, and in this way succeeded in synthesising a number of carbon compounds including amino-acids. There is therefore experimental evidence for this theory that the bricks out of which the building represented by a living organism is built could have been synthesised naturally at the surface of the young Earth.

cycle will therefore accelerate, compensating (and more than compensating) for the progressive exhaustion of the stocks of hydrogen.

The energy radiated by the stars, therefore, far from falling off with the passage of time, should actually increase steadily until the point is reached when the supply of hydrogen is almost completely exhausted. The following table (Russell, 1942) shows the probable evolution, past and future, of a star like the Sun, assuming that its present composition by mass is 51% hydrogen, 42% helium, 1% carbon and 6% other elements:

% HYDROGEN	SPECTROSCOPIC TYPE	ABSOLUTE MAGNITUDE	PERIOD (10^9 years)	EPOCH: + PAST − FUTURE (10^9 years)
91	*d*K8	+ 8.2	(39)	
81	*d*K5	+ 7.1	54	− 84
71	*d*K3	+ 6.2	33	− 41
61	*d*G8	+ 5.4	19	− 15
51	*d*G2	+ 4.7	10	0
41	F4	+ 3.9	5.2	+ 7.6
31	A8	+ 3.2	2.5	+ 11.4
21	A3	+ 2.6	1.1	+ 13.2
11	B9	+ 2.1	0.4	+ 13.9

In the fourth column are tabulated the periods required for the hydrogen content to change from $(x+5)\%$ to $(x-5)\%$ — e.g. from 56% to 46% at the present time — except in the top line (91% to 86%). The last column lists the corresponding epochs, past and future.

During the next 10^{10} years, therefore, the Sun should evolve into a white, high-temperature star which, although of type A, will be markedly less luminous than present-day type A stars (absolute magnitude about 0 or + 1). As we have already seen, this conclusion seems to constitute a very clear indication that the age of the Universe does not exceed several 10^9 years, for if it had been in existence for longer we should be able to observe stars that had already reached this stage — and in point of fact, we cannot. It can also be seen from the table that the Sun has not evolved to any considerable extent since the time of its birth, and that 3 or 4 thousand million years ago it must have been very much the same as it is today.

From this point of view the mass-luminosity relationship (Chapters 14, 15) constitutes still another argument in favour of the short time-scale, for if the stars had already been evolving for several 10^{10} or 10^{11} years we ought to be able to observe stars of widely different masses with the same luminosity, and vice versa; yet for the most part, as we have seen, they adhere closely to the mass-luminosity relation — with the sole exception of the white dwarfs, which will be discussed later.

In the same way one can investigate the future evolution of other types of dwarf, which, like the Sun, depend on the carbon-nitrogen cycle for their supplies of energy. It has been found that the future of these stars (which is much more extensive than their past) will be a slow evolution in the general direction of the Main Sequence. Having regard to the relative shortness of the span of past history, it can be seen why they cluster so closely about this region of the Hertzsprung-Russell diagram.

Whilst we are on this topic, it might be mentioned that the discovery of the sub-dwarf branch (Chapter 14) and of certain stars which do not observe the mass-luminosity relationship, suggests that the stars may not all have started life with the same supply of hydrogen — possibly because different conditions attended their birth — for we know that the luminosity of a star is primarily determined by the proportion of hydrogen that it contains.

We must now consider two other categories of star whose evolution appears to be very different from that of the majority: the red supergiants and the white dwarfs. And we shall see that, although they are wholly dissimilar in appearance, they may well have certain features in common.

The luminosities of the supergiants are so great (thousands of times that of the Sun, and often more) that they must rapidly exhaust their supplies of hydrogen. Whereas the Sun's energy output amounts to a mere 2 ergs/gm./sec., that of Y Cygni (which is 17 times more massive than the Sun but 30,000 times as luminous) is 3,600 ergs/gm./sec., and for the components of VV Cephei the figure rises to 17,000 ergs/gm./sec. It can be calculated that under these circumstances the supergiants must exhaust, not only their hydrogen, but the whole of their nuclear resources, in the course of several 10^{10} (or at most, 10^{11}) years. Now the Universe is certainly older than this, so we must assume that as the oldest supergiants die out, newly formed ones are continuously appearing.

The last 20 years' theoretical investigations of the evolution of the dark galactic nebulae, supported by the discovery of the galactic 'globules' (Chapter 16), appear to have established the fact that supergiants can slowly condense in sufficiently dense interstellar clouds, under the influence of their own attraction and of the light pressure of the stars around them; the process probably takes something like 10^8 years. A condensation of this sort would be, initially, a gigantic globule of hyperrarefied gas; progressive contraction would liberate more and more of its potential gravitational energy, and its temperature would slowly rise. Eventually it would start to radiate — at first in the infra-red, and then in the visible range. Gamow and Teller showed in 1940 that when the central temperature of the newborn supergiant reaches about 400,000°, the thermal agitation of its atoms will be sufficiently violent to unleash the first nuclear reactions — those involving protons and the nuclei of some of the lightest elements (deuterium, lithium, beryllium and boron). The first to start will be that between protons and deuterium nuclei, and this will continue until all the deuterium has been used up. Whilst the star's equilibrium is being maintained by the release of nuclear energy, contraction will be halted, though it is quite possible that irregular fluctuations might occur — as has been suggested by Gamow in proposing the theory that long period variables are stars at this stage of evolution. With the exhaustion of the deuterium supply, contraction will once again set in, and will continue until the central temperature has risen to 2 or 3 million degrees; lithium will then begin to react, followed by beryllium; during this phase, the star may become a classical Cepheid. Finally, after a further period of contraction, with ever-increasing temperature, it will be the turn of boron to start disintegrating. The star will thus pass through a series of alternating stages (represented diagrammatically in Fig. 727), in the course of which the light elements are used up in the reaction described in Chapter 15, and the star moves ever nearer to the Main Sequence. Here, at last, the carbon-nitrogen cycle will be able to get under way, and will make rapid inroads into the star's supply of hydrogen — which the preceding reactions left practically untouched, judging by the microscopic relative abundances of the light elements involved (something of the order of 10^{-6} times those of carbon and nitrogen throughout the whole Universe).

After passing rapidly through types A and B into O, a progress that might take no more than a few 10^7 years, the supergiant would find itself denuded of its hydrogen reserves. What will happen then? In theory, the same as before, whenever a particular nuclear fuel is exhausted:

Fig. 728 — GALACTIC GLOBULES PROJECTED AGAINST THE DIFFUSE NEBULA NGC 281 IN CASSIOPEIA. The straight line crossing the upper part of the plate is the trail of a meteor which traversed the field of the telescope during the exposure.
(Photo, Mount Wilson Observatory, 60-inch reflector, exposure 3h 30m; 1921)

Fig. 729 — DARK GLOBULES SEEN AGAINST THE DIFFUSE NEBULA M 8 IN SAGITTARIUS. According to B. J. Bok, these galactic globules are supergiants in process of condensation and are not yet luminous. *(Photo, Mount Wilson Observatory, 60-inch reflector, exposure 3h; 1919)*

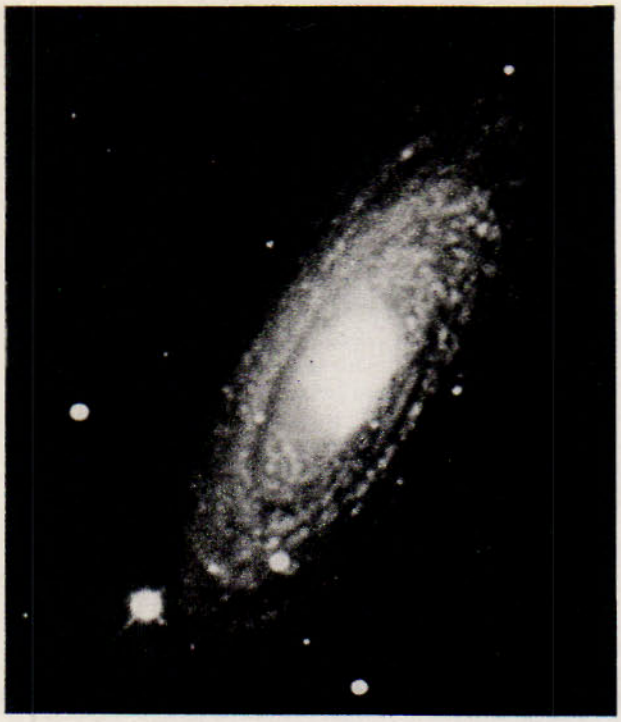


Fig. 730 — TYPE *Sb* SPIRAL NEBULA IN URSA MAJOR (NGC 2841). According to G. Lemaître, an object of this sort would result from an asymmetric collision of two primal clouds. *(Photo, F. G. Pease, Mount Wilson Observatory, 60-inch reflector, exposure 2h; 1912)*

Fig. 731 — A CURIOUS IRREGULAR NEBULA, NGC 2976 IN URSA MAJOR. According to G. Lemaître, an object of this sort would result from a central collision of two primal clouds. *(Photo, F. G. Pease, Mount Wilson Observatory, 60-inch reflector, exposure 3h; 1912)*

contraction. In the past, the increase of temperature resultant on contraction soon set off another reaction, the energy from which not only put a stop to the contraction but (as we saw when discussing the rôle of radiation pressure in maintaining the equilibrium of the most massive stars) contributed very largely to keeping the star at its current size. But now, with all the hydrogen exhausted, no further nuclear reactions are available to arrest the star's progressive contraction.

This, then, is the fate that will ultimately overtake all the stars, for sooner or later they will find themselves involved in the carbon-nitrogen cycle: it would seem that none is exempt, for the Main Sequence dwarfs are already in its grip, while the giants (as just described) are rushing towards the same doom.

Let us therefore consider, in general terms, what may be foreseen of the future of a star whose resources of nuclear energy are completely exhausted; we shall find that this unexpectedly leads straight to the white dwarfs and to the abnormal conditions and phenomena associated with them. The whole question has been investigated by, in particular, Chandrasekhar and his followers, and involves the concept of degenerate matter, whose essential properties have already been described in Chapter 15. Consider, then, the case of a star which has used up its last available gram of hydrogen: contraction will set in, but will be incapable of maintaining the very high rate of energy expenditure that characterised its final thermonuclear phase. The star will not only fall into a progressive contraction but will emit less and less radiation; it will accordingly develop very rapidly towards a new state of equilibrium. But during this period of contraction the mass of the star will not alter materially, and its density will therefore steadily increase; the process will ultimately be stopped by the attainment of a sufficiently high central pressure to produce the collapse of atomic structure: the core will suddenly relapse into the degenerate state, and the star will become a white dwarf.

It has frequently been argued during recent years, that the present existence of white dwarfs cannot easily be fitted into the framework provided by the short time-scale, for the Main Sequence dwarfs require several 10^{10} years to reach this stage of their evolution: in other words, the white dwarfs seem to be too old to have any place in a Universe so young. But this difficulty vanishes if we assume, as seems quite likely, that the white dwarfs are the remnants of rapidly evolving supergiants which

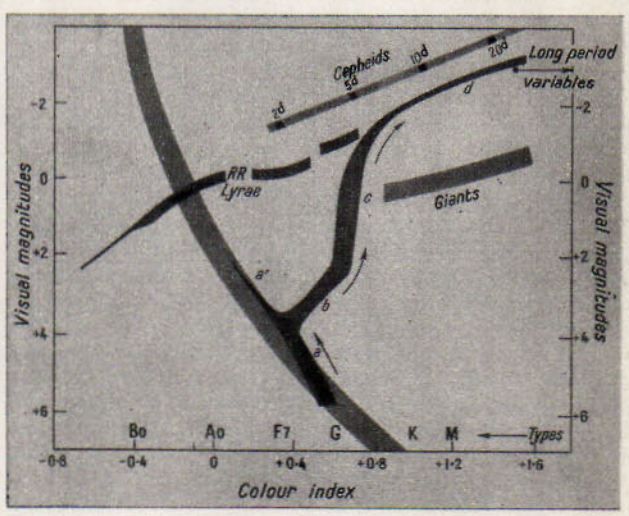


Fig. 732 — RUSSELL DIAGRAM FOR STARS OF POPULATION II (IN BLACK) WITH ITS RR LYRAE VARIABLES, AND POPULATION I (SHADED) WITH ITS CEPHEIDS

have reached the end of their thermonuclear phase and are already committed to the final, rapid contraction during which their luminosities decrease by 10 to 15 magnitudes. This contraction may not always, however, have time to liberate enough gravitational energy to restore equilibrium. Sudden halts might well be caused from time to time by the onset of degeneration, the release of the excess energy being accompanied by the simultaneous expulsion of some of the matter of the star. If this occurred suddenly enough it would produce a temporary brightening of the star, with the creation of a nova; furthermore, the process could recur at shorter or longer intervals until the final re-establishment of equilibrium, thus giving rise to recurrent novae. If, on the other hand, the ejection of matter and surplus energy took place relatively slowly and continuously, the star would develop an extensive envelope — as seems to happen in the case of the Wolf-Rayet stars. The very rarity of these stars is an indication that they represent a transitory and very brief stage in stellar evolution. Such envelopes may eventually develop into planetary nebulae — whose central stars, it will be recalled, are related to the Wolf-Rayets.

The precise form that the evolution of a star will take throughout this stage of contraction is in fact determined primarily by its original mass. Chandrasekhar has shown that degeneration can only supervene in a star which is effectively without hydrogen if its mass is less than about 1.5 times that of the Sun; the majority of the stars (the Sun, naturally, included) fall into this class. If, on the other hand, mass is more than about 1.7 times the Sun's, it will most probably become unstable during its contraction: it will either fly into fragments under the stress of increasing centrifugal force resulting from the progressively accelerating rotation that must accompany contraction (one of the currently envisaged causes of novae), or excess matter will be ejected in the more regular manner which is exemplified by the Wolf-Rayet stars, till finally its mass has fallen below the critical limit. Theory suggests that in the first group, contraction could continue indefinitely without equilibrium being attained, since the Fermi gas pressure would be insufficient (considering the mass involved) to support the weight of the surface layers. The Russian physicist, Landau, has shown that the star would therefore continue to shrink until its central matter is reduced to the nuclear state (already described in connection with the initial state of the Universe); it would then be a single, hyperdense nucleus, no more than a few tens or hundreds of miles in diameter, incapable of further contraction. If, to consider the remaining case, the star's mass were between these two limits (1.5 and 1.7 times that of the Sun) — and perhaps also if it were more than 1.7 and if, despite this, the star could avoid disintegration — a catastrophic development might occur during the phase of contraction; in this development Zwicky sees a possible explanation of supernovae. Such a star would, indeed, develop a degenerate nucleus (partial degeneracy); but under the ever-increasing pressure imposed by progressive contraction, a time might arrive when the inward-acting pressure of the surface layers would overcome the resistance of Fermi's gas: protons, and perhaps whole nuclei, would then be so tightly packed with free electrons that they would suddenly combine, with the formation of neutrons and possibly of a small central core of matter in the nuclear state. Because this matter is electrically neutral, the star would suddenly find itself deprived of the support of the enormous forces of electrical repulsion which had previously been supplied by the stability of its atoms and the individual existence of their particles. It would immediately collapse, liberating within a few hours its immense reserves of potential gravitational energy. This energy would escape almost instantaneously, disrupting the star's outer layers, which would fly into fragments expelled at velocities of the order of those observed in supernovae.

The fact that this development could only occur as the result of a very exceptional combination of circumstances agrees well with the observed rarity of supernovae, as compared with stars illustrating other phases of stellar evolution.

The existence of two distinct types of stellar population. — It is now many years since Shapley first drew attention to the absence of blue supergiants, classical Cepheids and cosmic gas in the globular clusters. The most luminous stars in these clusters are red giants, whose absolute visual magnitude does not exceed $Mpv = -3$. The cluster population thus appears to be different from that of the Sun's neighbourhood (which is represented in the familiar Russell diagram).

The significance of this observation was only properly appreciated when Baade, with the help of the great Californian reflectors, showed that the populations of the ellipsoidal galaxies and of the nuclei of the spirals (including that of the Galaxy) are identical with that of the globular clusters. This type of stellar population is known as Population II. Even more generally, Population II is only encountered between the arms of the galaxies — including again, the Galaxy.

Studies of the rotation of the Galaxy, and of the Sun's motion towards the apex, have produced highly discordant values of the latter when based on certain categories of object which include the globular clusters, the short period (RR Lyrae) Cepheids, and certain red giants with emission spectra of type *Me*. The distribution of the velocities of these objects relative to a local frame of reference shows that they constitute what is known as an asymmetrical current, and that although they appear to be high-velocity stars they are in fact extremely slow-moving, and are being rapidly overtaken by the Sun in the course of its rotation about the galactic centre. It is permissible to regard these stars as originating in the central regions of the galactic system; perturbations having made their orbits strongly elliptical, they are able to penetrate to peripheral regions such as the Sun's vicinity; but they are, so to speak, no more than casual visitors, which do not partake of the rotational motion of the galactic arm to which the Sun belongs.

We have already emphasised the seriousness of the error (only realised in fairly recent years) of assuming that a single period-luminosity curve could accommodate all the Cepheid-type variables, irrespective of their period — the RR Lyrae Cepheids as well as the classical Cepheids. The curve established by Miss Leavitt and Shapley was, as already explained, based on observations of classical Cepheids in the Magellanic Clouds, while its calibration was based exclusively on observations of RR Lyrae variables within the Galaxy. It is now firmly established that the two types of light curve refer to stars belonging to different populations.

It is assumed that Population II is the primeval stellar population of the Universe, whose original giants and supergiants — as well (as we have seen) as its sufficiently massive Main Sequence stars — have already disappeared.

Fig. 732 shows the diagram, analogous to the classical Russell diagram but referring to stars of Population II, which the Mount Palomar astronomers have derived from the study of the populations of the globular clusters. It has little in common with the classical diagram except the basic Main Sequence. Furthermore, the population

of the Sun's vicinity is mixed, and the members of the two populations can only be distinguished from one another by their kinematic characteristics (and, possibly, by their different chemical composition, of which more will be said later).

Evolution of Population II stars. — Along branch *a* (Fig. 732) the hydrogen of a star's core is being used up in thermonuclear reactions of the proton-proton or Bethe cycle. Its total luminosity and temperature are rising.

If the star is animated by a very rapid rotation it will continue along the Main Sequence towards *a'*. But the destiny of the much more numerous stars with slow rotations is very different. According to Schwarzschild's theory, the ordinary nuclear reactions will continue as long as the core of pure helium (now isothermal) does not account for more than 12% of the star's mass. When this stage is reached, Schwarzschild predicts, a reversal will set in: the isothermal core will contract and the outer envelope will expand, the star transforming into a red giant. These theoretical predictions are strikingly confirmed by the existence of the *b* branch of the Population II diagram. According to theory, stars with masses 1.5 times that of the Sun will take 3.5×10^9 years to pass along *b*. Stars whose masses are no greater than the Sun's will still be on branch *a*. (Initial bolometric magnitude $M_b > 4.5$). Branch *b* stars will, naturally, no longer conform with the mass-luminosity relation.

According to Sandage we should at the present time find branch *b* stars with bolometric magnitudes M_b between 4.5 and 4.0, the most luminous occupying the upper right-hand part of the diagram. Then, as the reserves of hydrogen are exhausted and the flow of nuclear energy is reduced, contraction will set in and the star will relapse along the horizontal branch towards the RR Lyrae variables and the 'blue dwarfs'. The end of this development may be marked by the occurrence of nova or supernova phenomena, degeneration into a white dwarf, etc.

In other words the Population II diagram appears to be a picture of the present distribution of a stellar population that was born some 4×10^9 years ago; at that time its stars occupied the lower end of the Main Sequence, with $M_b > 7$. Stars of this original population that were born with $M_b < 4$ will have disappeared, their evolution having been run through at a speed that is proportional to their luminosity at birth. Their disappearance would have been brought about by the cataclysmic dispersal of their material, by degeneration into the white dwarf condition, by fission, or by simple extinction.

Population II is thus the fundamental population of the Universe. Population I is no more than a comparatively recent epiphenomenon: its stars are built up of cosmic gas which, whenever it occurs, outlines the arms of the spiral nebulae — no gas, no arms, no Population I. Population II is widely distributed throughout the median planes of the spiral nebulae (M 31, for example, or our own), Population I being superimposed on it in the arms.

Distinguishing characteristics of Populations I and II. — The evolution of Population I stars would be comparable with that of the members of Population II; the former, however, are not all the same age, having been born during every epoch since the formation of the galaxies. Thus a dwarf like the Sun would be about 4×10^9 years old, whereas the Main Sequence giants and supergiants were born comparatively recently and will die out very rapidly; others, in the meantime, will continue to be born until the supply of cosmic gas is exhausted. In our own neighbourhood of space, stars of both populations are intermingled, and it is not always possible to distinguish between them.

Galaxies which at one fell swoop converted all their gas into Population II stars* will be unable to give birth either to spiral arms or to a Population I.

The two populations likewise differ in their kinematic characteristics. Population I stars, which are born in the arms, are endowed with 'circular' velocities, and are narrowly restricted to the galactic plane. Population II stars have, in general, very much higher velocities, which lie, if not in completely random directions, at least in directions less rigidly confined to the median plane than those of Population I. Their orbits are elliptical, and they are sometimes to be found in high galactic latitudes.

There is even a tendency nowadays to group different categories of object together into galactic sub-systems on a basis of their relatively ellipsoidal or flattened spatial distributions. This is, perhaps, an abuse of 'selectivity', but the near future may be expected to decide the question.

The members of the two populations are also distinguished from one another by a fairly marked dissimilarity of chemical composition. In particular, the contemporary Population II is ten times less rich in metals (by mass) than Population I. It looks as though the Population II stars were born from the primeval gas before stars appeared in any considerable numbers.

The formation of heavy elements in the Universe. — This is one of the principal enigmas that remain unsolved at the present time. Since, under the conditions obtaining at the centre of a normal star (20 million degrees, 9×10^9 atmospheres), the thermonuclear reactions result only in the production of helium from hydrogen, with the annihilation of the light elements under the catalytic action of carbon, the creation of the heavier elements would involve (in view of the exceptional stability of the helium nucleus) considerably higher temperatures. Chandrasekhar, basing his calculations on the proportions known to exist today, deduced a temperature of 8×10^9 degrees and a density of the order of 10^7 gm./cm^3. It is likely, however, that lower temperatures than this would suffice if even higher densities could be invoked.

Is it possible to envisage a hyperdense pre-stellar phase? Such a state is compatible with the theories of expansion, and would characterise the epoch preceding the onset of the expansion — Lemaître's theory of a 'primeval atom'. But this hypothesis is difficult to reconcile with the existence of two different stellar populations, and, in particular, with their very dissimilar proportions of heavy elements.

At the present time it seems more reasonable to suppose that these heavy elements were formed in the cores of particularly massive Population II stars. (Calculation shows that this would be rendered possible by contraction of the core.) Such stars rapidly exhaust their sources of nuclear energy, and under certain circumstances might become unstable to the point of experiencing a supernova explosion, which would serve to scatter he heavy elements throughout space.

The old Population I stars, such as the Sun, would have been formed after this original distribution of the heavy elements. Once the hypermassive Population II stars had all disappeared, no further distribution of metals could occur. Population I supergiants would, however, continue to condense from the clouds of cosmic gas as long as any remained.

Having now outlined the principal stages in the life-history of stars, and their relations one to another, we

* A modern hypothesis suggests that intergalactic collisions may account for the absence of gas in galaxies of this kind.

must now enquire whether the life of a galaxy can be similarly analysed.

Evolution of the galaxies. — Despite the quantity of work which has been done in this field — notably by B. Lindblad and his followers — we are still far from arriving at any very clear account of galactic evolution.

Lindblad, following Jeans, supposes that soon after their formation the galaxies were in the form of spherical or slightly flattened elliptical nebulae; from these nuclei spiral arms were later to develop, under the influence of external perturbations which modified the orbits of the stars within the nebulae, causing them to leave the central mass. As a result of this ejection of matter, the system would become increasingly flattened, a process that would continue until a value of about 0.6 was attained; the ejection of matter would then accelerate, since the orbits nearest the edge of the central mass would become unstable, and the expelled stars would form themselves into rings (barred spirals) or spiral arms (normal spirals) according to the particular circumstances of their ejection; the various forms assumed by the arms would be determined by the relation between their total mass and that of the nucleus. According to this view, the spiral arms represent the actual trajectories described by the stars ejected from the nucleus, so that the direction of the nebula's rotation is that tending to unwind the arms; these therefore precede the nucleus.

We have seen in Chapter 16, however, that this conclusion has not been incontestably established, and that many astronomers champion rotation in the opposite sense. Then again, we have seen that photometric measures have proved the elliptical and the spiral nebulae to be of much the same size, in which case it is difficult to see how the former can be the nuclei of the latter. Finally, this scheme of evolution does not seem to fit very easily into the short time-scale, nor do the mean ages of elliptical and spiral nebulae appear to be so very different from one another.

We are thus forced to conclude that the galaxies, like the stars, acquired their individual characteristics during the decelerating phase of the Universe's expansion, and have only developed to a very limited extent since that time. This leads us to the Abbé Lemaître's account of the evolution of the galaxies. Lemaître supposed that during the epoch of deceleration, fluctuations of density (due to random motions, or to the collisions of the gaseous or stellar clouds) would confer on the resultant medium a density whose value exceeded the critical value at which equilibrium was possible. Attraction would therefore overcome repulsion, and individual stable galaxies would be born; later, their separations would be increased enormously by the expansion of space.

Assuming this, how can the development of spiral arms be explained? Lemaître suggests that rotation sufficiently rapid to cause the ejection of arms could have resulted from face-to-face but eccentric collisions between individual clouds; another possible causative mechanism is one analogous to that whereby Moulton explained the tidal extraction of a filament of matter from the Sun (p. 436). Fig. 723 shows that spiral arms could result from the near encounter of two galaxies (which during the decelerating phase would have been very close to one another), for the points of emission of the matter on the periphery of the two galaxies, or of their central nuclei, would be progressively displaced in the direction of the perturbing body. Given these conditions, the arms would mark the positions simultaneously occupied by stars which had been ejected from different points on the periphery (a situation reminiscent of the curved tails of comets). They would thus revolve in the winding-up sense, i.e. would follow the nucleus — as advocated by Hubble.

In this way the simultaneous existence of spiral and elliptical nebulae can be accounted for. The irregular, Magellanic-type galaxies were perhaps the result of rare, centre-to-centre encounters between galaxies, which would produce no rotation and therefore set up no rotational symmetry within the system.

Whatever the final verdict may be, as between these two different theories, it seems unlikely that the arms of the spiral nebulae can have made a very large number of revolutions since the time of their formation, since the varied perturbations to which they are subjected would tend to break them up rather quickly. After some dozens of revolutions, which many spirals must have made since their formation, one would therefore expect the spiral structure to have become considerable, irregular and broken-up; this is actually observed in the *Sc* and *SBc* spirals.

Our understanding of the dynamics of the galaxies is, however, still at too rudimentary a stage for us to be able to form any precise picture of their future evolution. Nevertheless we can, within the framework of the theories already described, predict fairly plausibly some aspects of the future of the Universe.

The future, and the end. — The characteristics of universal evolution that seem to be of the most fundamental significance are: the expansion of the Universe, and the irrevocable wastage of the reserves of energy in the interiors of the stars. Between them, these apparently irreversible processes determine the future evolution of the Universe.

The expansion is ceaselessly increasing the separation of the galaxies (whether isolated, or grouped into clusters), and we must envisage a time when the velocity of this mutual recession will exceed that of light. From then onwards, no galaxy will be able to receive the radiation emitted by any other; the Universe will be a hypersphere of ever-increasing size and ever-decreasing density, in which the nebulae, singly or in groups, will voyage alone into a limitless future. Each will be completely isolated from all the others; each will be situated at the centre of a sphere of space filled with its own radiation, and containing nothing else; even this will become ever more diluted.

Today the radius of the observable region is about 1.5×10^9 light years: and although this sphere is no more than a few hundredths of the total volume of the Universe, it is nevertheless large enough to contain some thousands of millions of galaxies. Before these cross the fatal frontier into eternal invisibility the Universe will have to expand by a factor of perhaps a hundred or a thousand — and this, at the present rate, will take several 10^{10} or 10^{11} years.

It is hardly necessary to add that no human being will still be on the scene to witness that event. Taking only a single factor into consideration — the Sun's radiation will long before then have dwindled to such an extent that the Earth's atmosphere will have condensed to liquid air, covering the eternal ice of the oceans to a depth of 30 feet — assuming, that is, that the slow infiltration of the oceans into the Earth's interior has not dried them up by then, uncovering, for the rest of eternity, a frozen crust indistinguishable from that of the continents.

There is, however, another possibility. Fifteen thousand million years from now the Sun, having exhausted its stock of hydrogen*, will relapse towards the status of

* Assuming that its present hydrogen content is roughly 50%; if it were only 30% this period would be reduced to about 10^{10} years.

a white dwarf, and it is possible that it will then pass through the nova stage: as we have seen, this is by no means unlikely during the unstable phase of contraction. In this event the Earth and all the other planets will be vaporised, or at any rate their outer layers would melt — the consequences for anything still living at that time would be equally catastrophic.

But whether the end is to be refrigeration or ignition, the future span of life on the Earth would appear to be limited to 10^{10} or 1.5×10^{11} years at most, even ignoring the possibility of numerous other catastrophes — astronomical and otherwise — that could wipe it out before then.

It would seem, then, that every star that now exists, and every supergiant yet to be born in the heart of the interstellar cloud during the millions of centuries ahead of us, is condemned to the ultimate collapse which will transform it into a degenerate white dwarf or a hyperdense nuclear dwarf. What will be the final end of stars that have reached this stage? Degenerate white dwarfs will no doubt be able to continue radiating with ever decreasing intensity for some 10^{10} years, before reaching the black dwarf state — becoming first infra-red, and then at last cold and inert for ever, despite the prodigious reserves of energy still locked up in their degenerate matter.

Stars of the second type, when at last they reach the nuclear state — by way, perhaps, of such vicissitudes as passage through a supernova stage — will similarly be immobilised forever, unless the atomic super-nucleus should be unstable, endowed with some form of radioactivity analogous to that which may have existed in the primal nucleus from which the Universe sprang.

But these ultimate hypotheses are speculative in the extreme, and it has to be admitted that despite the brilliant advances of recent years, the past and future history of the Universe still remains to a very large extent mysterious and unfathomed.

Thus we have come to the extreme limits of space and time that can be reached by present-day theory and observation, and we must now return once again to Earth in order to learn something about the instruments and observational techniques that the astronomer employs in his exploration of the heavens.

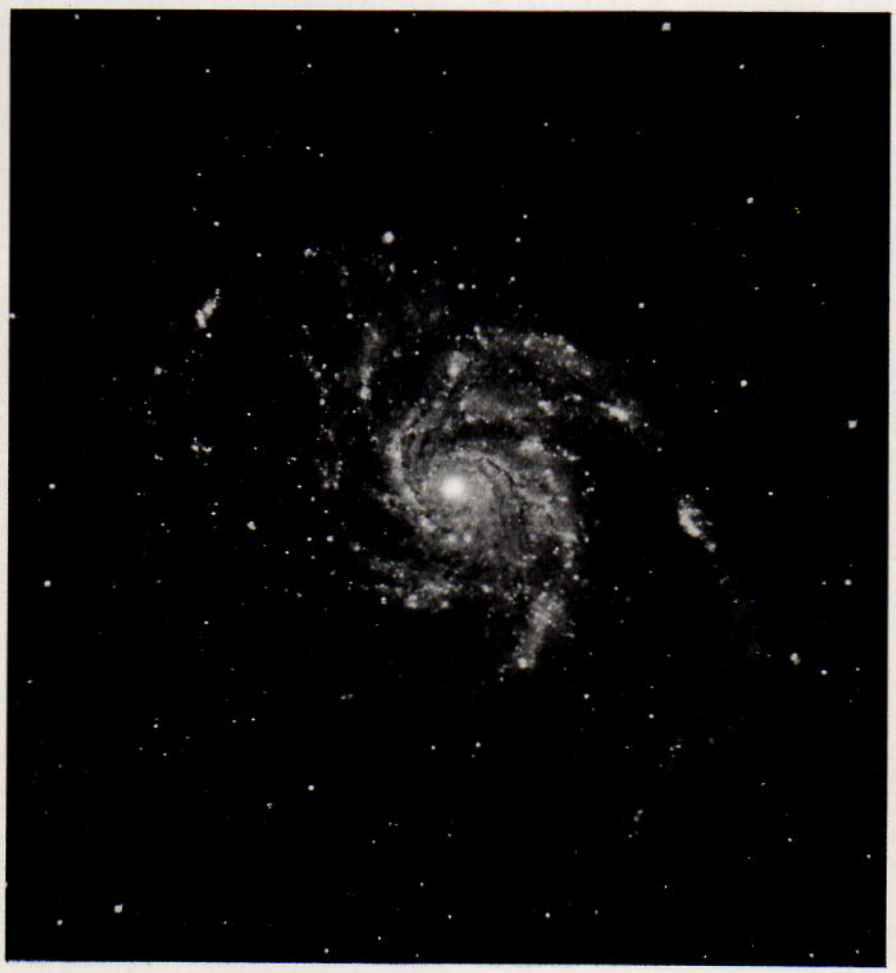


Fig. 733 — SPIRAL NEBULA (M 101) IN URSA MAJOR. The magnificent whirlpool structure of the nebulae raises the difficult problem of their origin *(Photo, Mount Wilson Observatory, 60-inch reflector, exposure 7h 30m; 1910)*

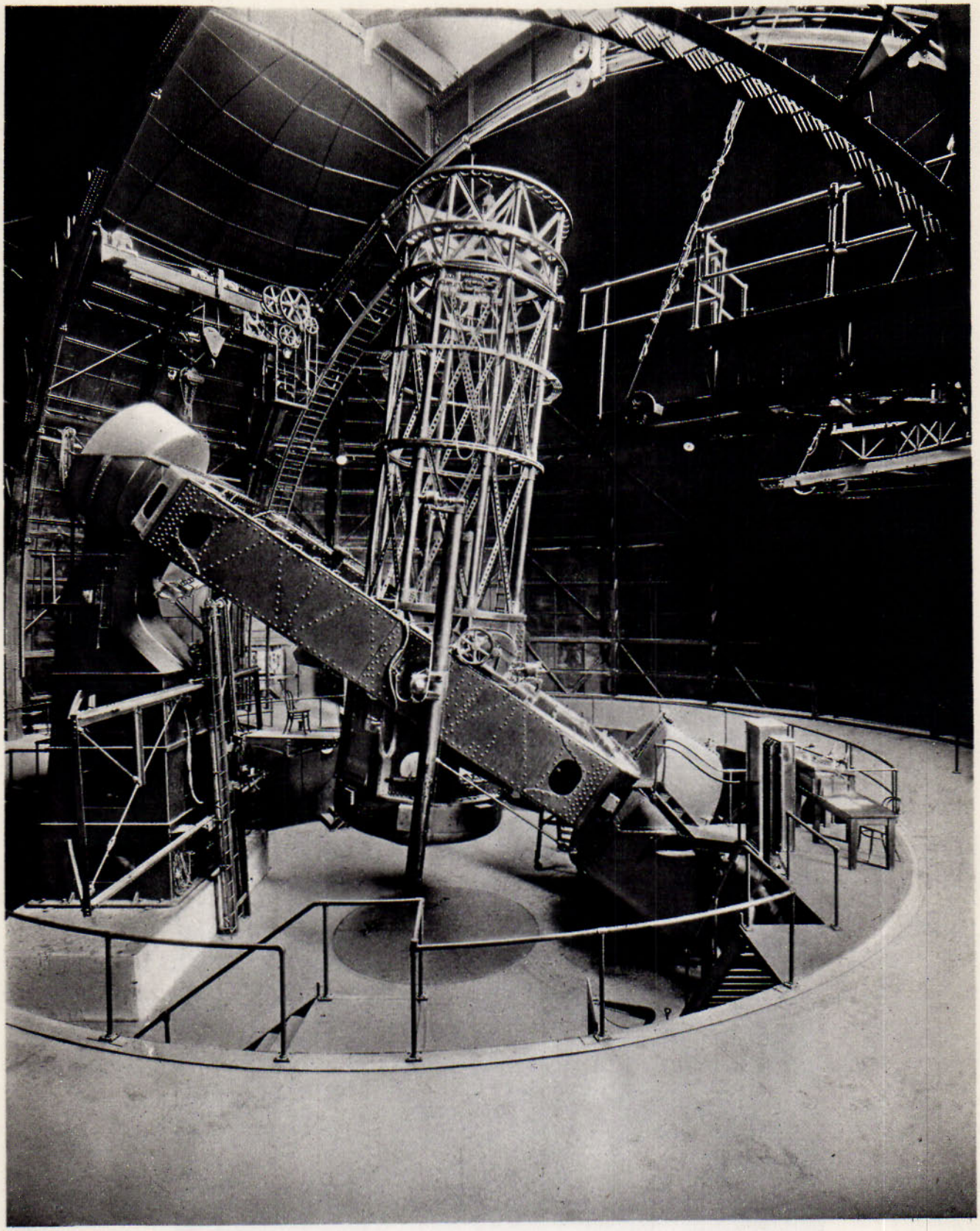


Fig. 734 — THE 100-INCH REFLECTOR OF THE CARNEGIE INSTITUTION AT MOUNT WILSON (CALIFORNIA), with which many of the greatest astronomical discoveries of the last 40 years have been made. *(Photo, Mount Wilson Observatory)*

BOOK IV

ASTRONOMICAL INSTRUMENTS AND TECHNIQUES

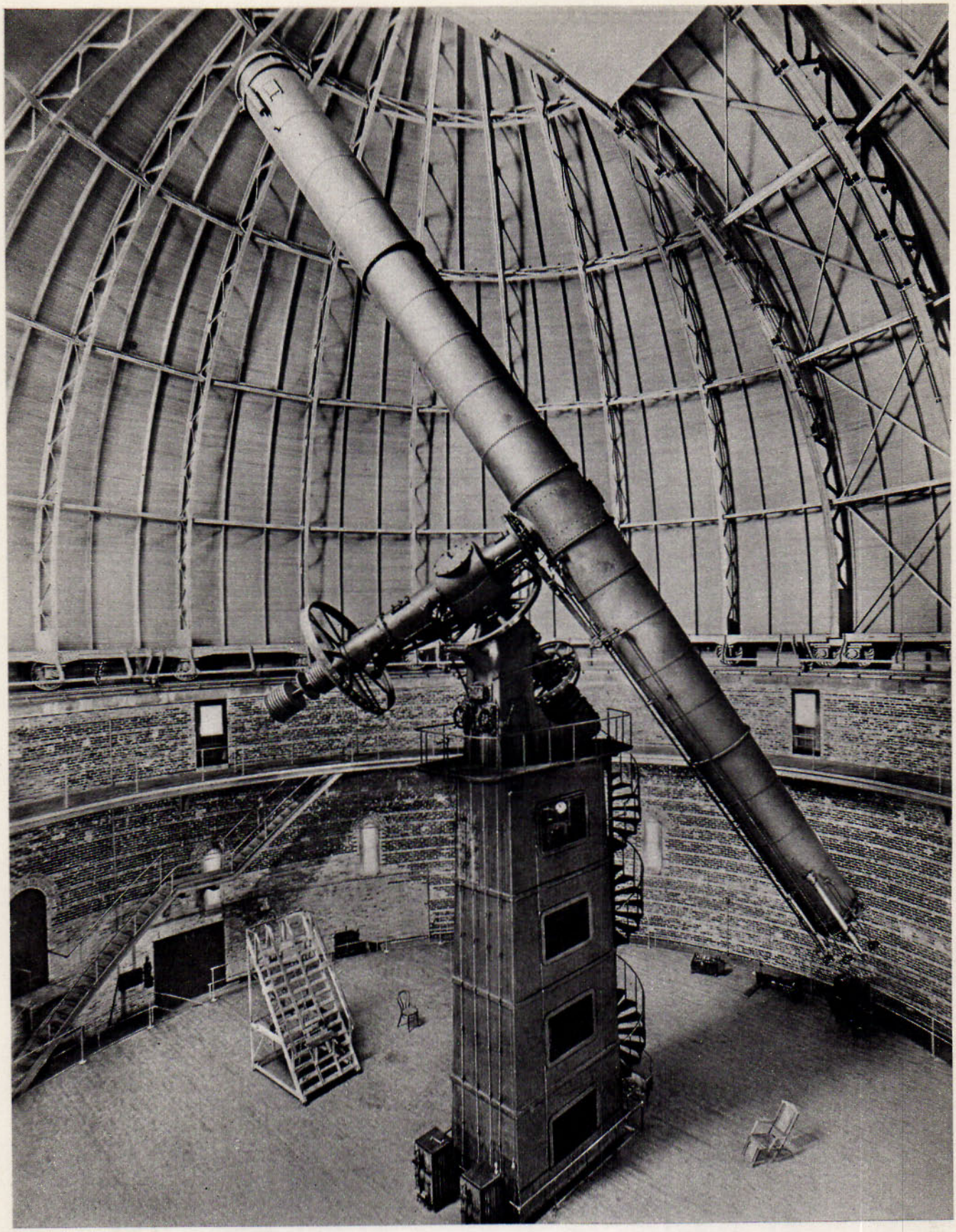


Fig. 735 – THE GREAT REFRACTOR OF 40 INCHES APERTURE AND 63 FEET FOCAL LENGTH AT YERKES OBSERVATORY, UNIVERSITY OF CHICAGO, WILLIAMS BAY, WISCONSIN

Fig. 736 — THE PIC DU MIDI OBSERVATORY, HAUTES-PYRÉNÉES, FRANCE. *(Photo, Alix)*

CHAPTER 19

INSTRUMENTS AND METHODS OF MEASUREMENT

In the preceding three Books we have been concerned with surveying present-day astronomical knowledge, and only the indispensable minimum has been said about the means and techniques which have provided this knowledge. The next two chapters will be devoted to the fundamental physical methods and the chief instruments that the astronomer employs.

Telescopes and observatories. — The first instrument to have transcended the limitations of man's natural eyesight was the astronomical telescope. It was invented — more or less by accident — in Holland, about 1609 — and was at once put to astronomical use by Galileo; this primitive optical arrangement is still known by his name. The astronomical telescope consists in principle of two glass lenses. One is of relatively large diameter and focal length, and is known as the objective since it is pointed towards the object that is being observed; it concentrates the light from this object, forming in its focal plane a real, inverted, aerial image. The second lens, of smaller diameter and shorter focal length, is termed the ocular or eyepiece since it is to this lens that the observer puts his eye; its function is to facilitate the examination of the primary image, formed by the objective, of which it gives a magnified virtual image (Fig. 737). In the Galilean telescope the ocular is a diverging (negative) lens; this combination of convex object glass and concave ocular is still employed in opera glasses, but in astronomical

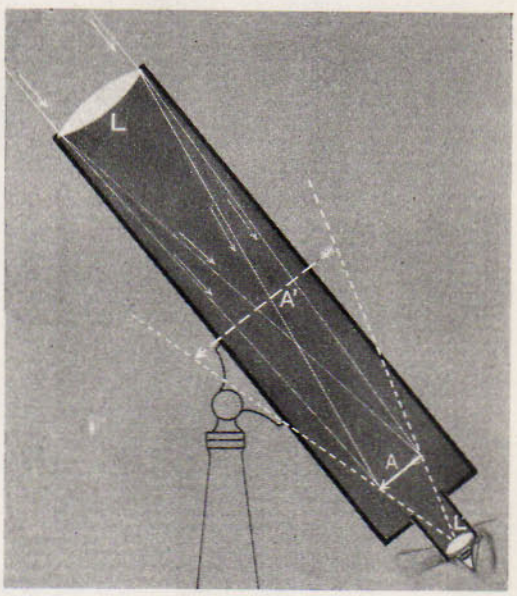


Fig. 737 — PRINCIPLE OF THE ASTRONOMICAL REFRACTOR. The objective L forms a real, inverted image of an infinitely distant object at A; the eyepiece L′ then forms a magnified virtual image of this at A.

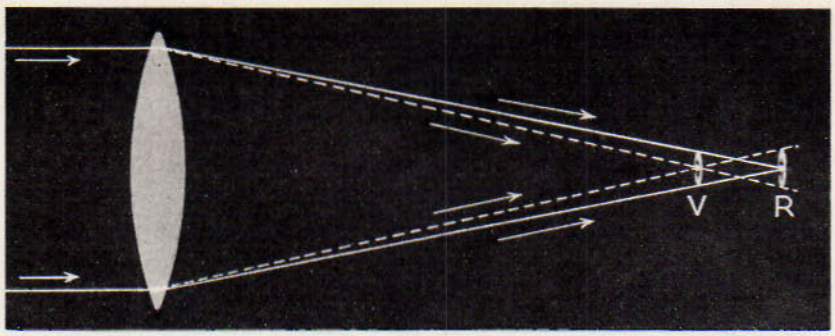


Fig. 738 — CHROMATIC ABERRATION. A simple lens focuses the violet rays to V and the red rays to R. The lines showing the course of the corresponding rays explain diagrammatically the formation of a fringe of colour round the image.

telescopes the negative ocular has long been replaced by a converging lens which thus acts as an ordinary magnifying glass, giving a wider field of view.

The early instruments of the seventeenth century were mostly of very poor optical quality, not only because optical glass was at that time very far from achieving the transparency, homogeneity and polish that it has today, but also because their objectives were simple lenses and were therefore affected by numerous aberrations — that is to say, they were incapable of bringing all the rays from a point object to precisely the same point in the focal plane.

One of the most troublesome of these aberrations is that known as chromatic aberration. A simple object glass does not refract radiations of different colours to the same extent, since the refractive index of any medium is a function of the wavelength of the radiation used to measure it. Thus blue light (short wavelength) is brought to a focus nearer to the objective than that of red light, while intermediate colours are brought to foci at intermediate distances between these two extremes (Fig. 738). Accordingly, if the image of a star, for instance, is accurately focused in yellow light (to which the eye is most sensitive) it will be found to consist of a central yellow disc surrounded by a violet or purplish halo, while the image of a planet will be bathed in a haze of unfocused blue and red light, much to the detriment of the clear definition of its surface details.

A partial remedy for this trouble is to make the focal length of the objective several hundred times greater than its aperture, and this fact was responsible for the construction of the enormously long and unwieldy refractors that are to be seen in seventeenth-century engravings.

The images produced by these cumbersome instruments were, even so, far from perfect, and it was not until the middle of the 18th century that the English optician J. Dollond discovered a method of constructing an achromatic objective by combining a converging lens of low dispersion, constructed from a light glass of the type known as crown, with a diverging element of heavier (flint) glass giving greater dispersion; the opposite chromatic aberrations of these two components could be

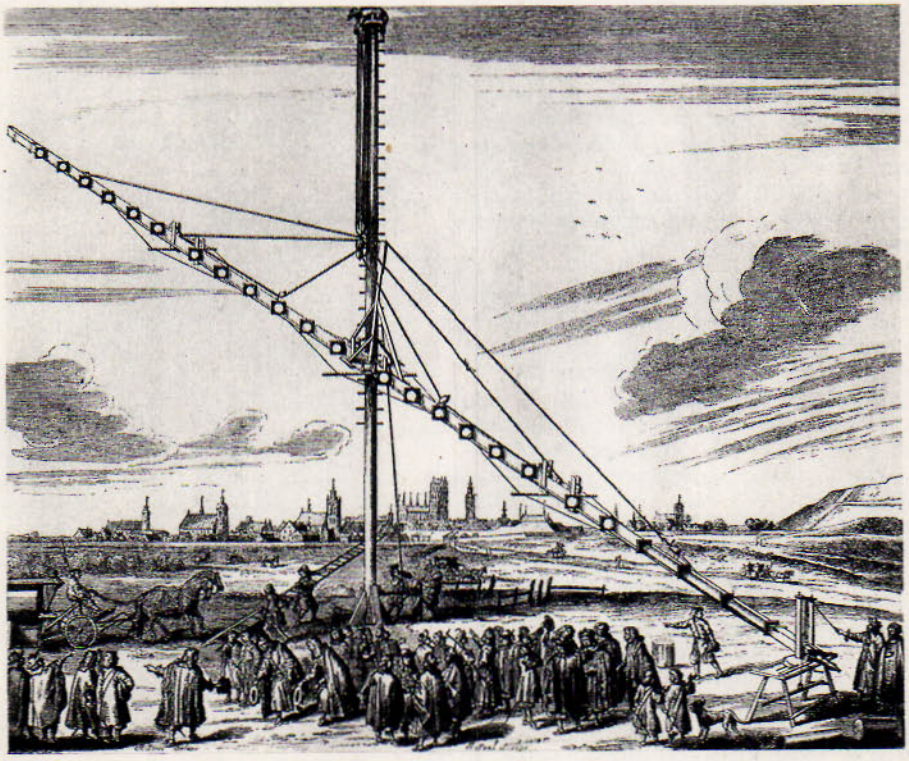


Fig. 739 — HEVELIUS'S MAMMOTH ASTRONOMICAL TELESCOPE. (Engraving from *Machina Coelestis*, Johann Hevelius, *c.* 1670)

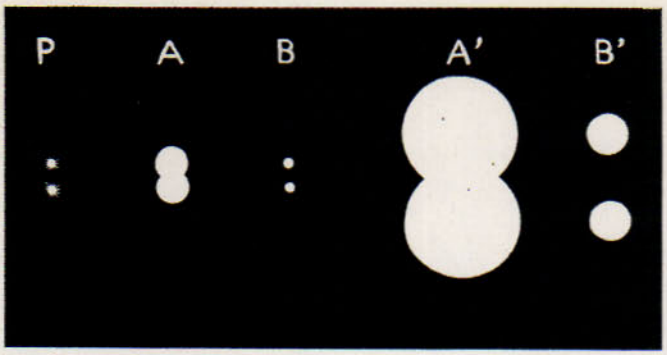


Fig. 740 — EXPLANATION OF RESOLVING POWER AS APPLIED TO THE CASE OF A DOUBLE STAR. A small instrument shows two closely adjacent points of light P, whose diffraction patterns overlap one another; they will continue to overlap, however great the magnification used (A, A'). An instrument of wider aperture, on the other hand, will produce smaller diffraction patterns and will therefore completely separate the two components (B, B').

adjusted to counteract one another. With this discovery it was possible to reduce the focal lengths of refracting telescopes to the more reasonable proportions which we are familiar with today, embodying focal rations (i.e. relation of focal length to diameter of object glass) generally in the region of about 15 or 20. Larger and larger refractors were constructed during the next 150 years or so, culminating in the great 40-inch telescope (63 feet focal length) which was put into operation at Yerkes Observatory, near Chicago, in 1897.

This trend in telescope design towards ever greater aperture was motivated by the desire not only to obtain brighter images — and hence be able to see fainter and fainter objects (see table, p. 304) — but also to increase the instrument's resolving power, i.e. its ability to reveal finer and finer detail on, for example, the surface of a planet.

The undulatory nature of light gives rise to a very important phenomenon known as diffraction. One of its effects is a lateral spreading of a ray of light when it encounters some obstacle in its path (shown, for instance, by the fringes at the edge of the shadow of a screen illuminated by a point source — see Fig. 368) or when it passes through an aperture, such as a telescope objective, which limits its diameter.

This latter effect, whose magnitude is inversely proportional to the diameter of the aperture, is shown in its pure form by the telescopic image of a star, which is seen to consist of a small central disc (known as the spurious disc), surrounded by a series of decreasingly bright rings. (See Fig. 605).

It can be appreciated that this will impose a limit on the resolving power of the instrument: if the angular separation of the components of a double star is too small, their two diffraction discs will overlap one another to such an extent that they will merge into one, and the double will not be resolvable with the instrument in question. Resolution can only be achieved by using an instrument of larger aperture: since it produces smaller spurious discs (Fig. 740) the overlapping will be reduced. Thus — contrary to what one might at first suppose — the larger the telescope, the smaller the stars appear.

A telescope's resolving power must therefore be clearly distinguished from its magnification, which depends solely upon the relation between the focal lengths of its objective (F) and its ocular (f): $M = F$. Simply by changing the ocular, one can change the magnification at will (from × 50 to × 2000, for instance), whereas the instrument's resolving power is fixed and unalterable, being determined by the clear aperture of the objective and the effective wavelength of the light — which is about 0.55 μ for visual observation.

The observation of double stars has shown that these will just be separated* (i.e. the two spurious discs just touching, but not overlapping) when the angular separation of their centres, R, is such that $R = \frac{5''}{D}$, D being the aperture in inches. The following table gives the theoretical visual resolving power of different apertures:

APERTURE (in.)	3	6	12	24	36	60
Resolving power	1.7″	0.8″	0.4″	0.2″	0.14″	0.08″

It should be mentioned that, more often than not, currents and eddies in the atmosphere produce a trembling and wandering of the images as a result of which the telescope's actual performance falls short of its theoretical resolving power. This trouble is encountered particularly at observatories situated near sea level, and for this reason many large instruments have been installed in mountain observatories. The Pic-du-Midi Observatory in the Pyrénées, for example, frequently enjoys outstandingly favourable atmospheric conditions, which have made it possible for fine photographs of the Moon and planets to be taken.

Owing to the technical difficulties encountered in the manufacture of very large blocks of homogeneous glass suitable for telescope objectives, no refractor larger than the Yerkes instrument has ever been constructed. This, however, is not the only reason why the construction of very large refractors has been suspended since the end of the 19th century, though it is the most important one. Other difficulties are associated with the use of refractors of really large aperture, the most important of which is chromatic aberration. This cannot be wholly suppressed but only very much reduced (to about 5% that of a simple, non-achromatic lens); with very large focal lengths (45 to 60 feet) the residual chromatic aberration becomes rather troublesome, resulting in a spread of the

* Assuming, of course, that sufficient magnification is used for all the detail in the image formed by the objective to be distinguished. The magnification just sufficient to ensure this is in the neighbourhood of 20 × D, the aperture in inches — anything from × 200 to × 1000 with instruments in common use.

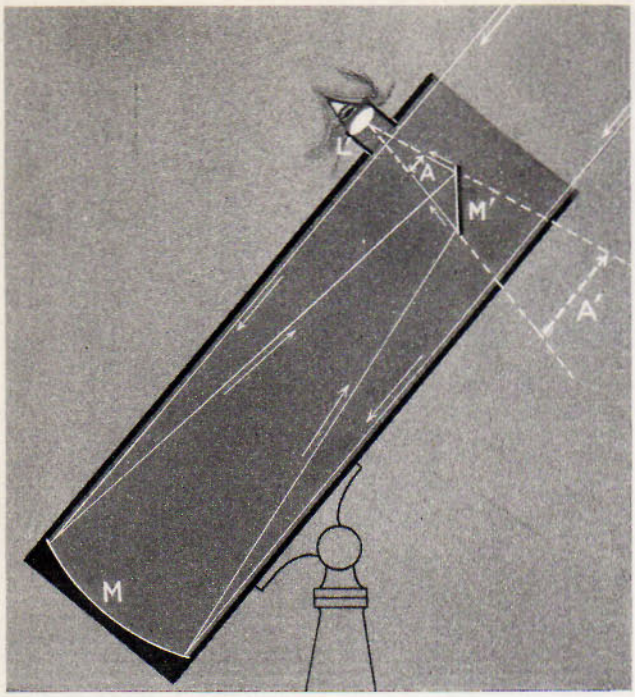


Fig. 741 — PRINCIPLE OF THE NEWTONIAN TELESCOPE. The mirror M produces, after reflection of the rays at the secondary mirror M′, a real image A of an object at infinity; the eyepiece L′ then forms a magnified virtual image of this A′.

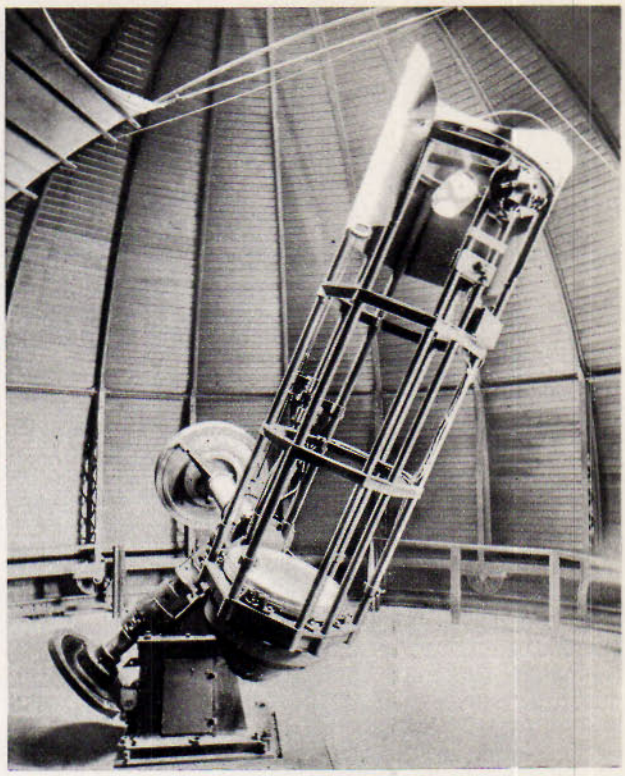


Fig. 742 — THE 24-INCH TELESCOPE AT YERKES OBSERVATORY. Compare with the diagram of the Newtonian telescope preceding.

focal 'plane' to a thickness of several millimetres, or even centimetres. Furthermore, the thickness of large object glasses is necessarily quite considerable — 8 inches or so — and at such apertures the light absorption in the glass is by no means negligible; increasing the aperture to obtain greater light grasp is no solution since the thickness of the lens must also be increased, which nullifies the hoped-for increase of light grasp. These disadvantages are increased if the image is to be recorded photographically, for the silver emulsion is more sensitive than the eye (which it has largely replaced during the past 50 years) to the blue and violet radiations.

The following table lists the 15 largest refractors at present in operation throughout the world:

THE WORLD'S LARGEST REFRACTORS
(Aperture > 25 inches)

OBSERVATORY	APERTURE (inches)	FOCAL LENGTH (feet)	DATE OF COMPLETION
Yerkes, Williams Bay (USA)	40	63	1897
Lick, Mt Hamilton (USA)	36	58	1888
Paris, Meudon (France)	33	53	1896
Potsdam (Germany)	32	39½ (*p*)	1905
Allegheny, Pittsburgh (USA)	30	46 (*p*)	1914
Bischoffsheim, Nice (France)	30	59	1887
Pulkova (USSR)	30	46	1886
Greenwich (England)	28	28	1894
Bloemfontein (S. Africa)	27	40	1928 (1)
Vienna (Austria)	27	34½	1878
Johannesburg (S. Africa)	26½	35	1926
Greenwich (England)	26	22	1897
McCormick, Charlottesville (USA)	26	32½	1873
Johannesburg (S. Africa)	26	36 (*p*)	1925 (2)
Washington (USA)	26	32½	1873

(1) Southern Station of the Observatory of the Universy of Michigan.
(2) Southern Station of the Observatory of Yale University.
(*p*) Photographic instrument.

Astronomers have therefore turned their attention to the reflecting telescope (also invented during the seventeenth century) whose objective is not a lens but a mirror. The rays entering a reflector are collected by a concave mirror, parabolic in section, which is mounted at the lower end of the tube; thence they are reflected back towards the top of the instrument as a convergent beam which comes to a focus in the mirror's focal plane. Here, at the so-called prime focus, it may be recorded photographically, but frequently it is reflected to a more accessible position at the side of the tube by means of a small, plane, secondary mirror. This arrangement of concave primary and flat secondary mirror was invented by Newton during the latter half of the seventeenth century (Fig. 741).

Another arrangement dating from the seventeenth century was that developed by Cassegrain, in which the small secondary mirror is not flat but a convex hyperboloid, and reflects the convergent beam back through a hole in the centre of the primary; here, at the back of the instrument, a magnified image is formed where it can conveniently be observed from a position similar to that occupied by the observer with a refractor.

The fundamental advantage of the reflector over the refractor is its complete freedom from chromatic aberration, since the laws of reflection are identical for all wavelengths. Another advantage, almost as important, is that since the light does not pass through the mirror this does not have to be made of highly transparent and homogeneous glass: its front surface, ground and polished to the requisite figure, alone plays any part in the formation of the image. Until about the middle of the nineteenth century the mirrors of reflectors were, indeed, made of a special bronze alloy, which could be made to take a sufficiently high polish.

Even Herschel's great telescopes, constructed towards

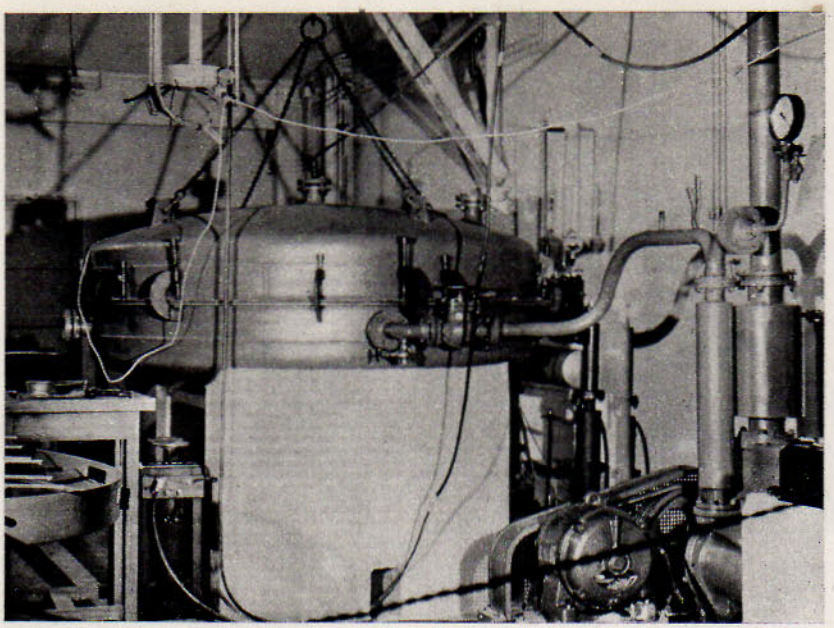


Fig. 743 – 82-INCH ALUMINISING CHAMBER FOR COATING LARGE ASTRONOMICAL MIRROR BY HIGH-VACUUM SUBLIMATION. Note on the right the vacuum pump and gauge. *(Institut d'Astrophysique de Paris. Photo, Maguery)*

the end of the eighteenth century (which culminated in his 48-inch reflector), had mirrors of this kind; so had the 72-inch constructed about the middle of the nineteenth century by William Parsons, fifth Earl of Roose, which remained the world's largest telescope until the advent of the Mount Wilson 100-inch, three-quarters of a century later.

Speculum metal was replaced by glass for telescope mirrors – glass being much easier to work – following Foucault's discovery in the mid-nineteenth century of a chemical method of depositing an even and highly reflective film of silver on a glass surface. Today even this method is being largely abandoned in favour of a more modern technique perfected by the American physicist J. Strong, some 25 years ago: pieces of metallic aluminium are heated electrically and evaporated in a vacuum, and a film of aluminium is deposited on the cleaned surface of the mirror. Aluminium has a double advantage over silver: the film is tougher and tarnishes much more slowly, lasting several years as against a few months; and it is considerably reflective in the ultra-violet, to which silver is relatively transparent and therefore only weakly reflective.

The fact that the mirror of a reflector does not have to be made of high-quality optical glass has opened the road to the construction of ever larger telescopes since the beginning of this century: this steady increase in aperture was necessitated by the astronomer's desire to study ever more distant objects – notably the extragalactic nebulae – and to submit their spectra to detailed analysis. Thus were constructed, successively, the 60-inch (1908) and the 100-inch (1917) at Mount Wilson, the 72-inch of the Dominion Observatory, Victoria, Canada (1919), the 74-inch at the Dunlap Observatory, Canada (1933) and the 82-inch at the McDonald Observatory (1939) in the United States; all these were eclipsed by the Mount Palomar 200-inch, whose planning and construction required 20 years of endeavour and the expenditure of more than $ 6,000,000.

The parabolic mirror has, however, one serious shortcoming: images situated far from the optical axis are affected by the aberration known as coma (since it transforms what should be a point image into a small, flared patch which is somewhat reminiscent of a comet). The magnitude of the aberration increases very rapidly with increasing distance from the axis of the instrument, and its effect is seriously to reduce the size of the usable field; that of the Mount Wilson 100-inch (relative aperture *f*/5), for example is barely large enough to cover the full Moon.

The following table lists the world's largest reflectors, either in operation or under construction:

THE WORLD'S LARGEST REFLECTORS
(Aperture > 59 inches)

OBSERVATORY	APERTURE (in.)	FOCAL LENGTH (feet)	DATE OF COMPLETION
(Soviet Union)	236	–	projected
Mount Palomar (USA)	200	55	1948
Lick, Mount Hamilton (USA)	120	50	1958
(South Africa)	120	–	projected
Crimea (Soviet Union)	103	(60)	under construction
Mount Wilson (USA)	100	42	1917
Herstmonceux (England)	98	–	projected; work suspended
McDonald, Mount Locke (USA)	82	–	1939
Kitts Peak (USA)	80?	–	projected
(Egypt)	74	–	complete; not yet erected
Haute-Provence (France)	74	–	1958
Radcliffe, Pretoria (S. Africa)	74	–	1958
Mount Stromlo (Australia)	74	–	1955
Dunlap, Toronto (Canada)	74	30	1935
Dominion, Victoria (Canada)	72	30	1919
Perkins, Delaware (USA)	70	25	1932
Harvard, Oak Ridge (USA)	61	26	1937
Harvard, Bloemfontein (S. Africa)	60	26½	1933
Mount Wilson (USA)	60	25	1908
Bosque Alegre (Argentina)	60	25	1942

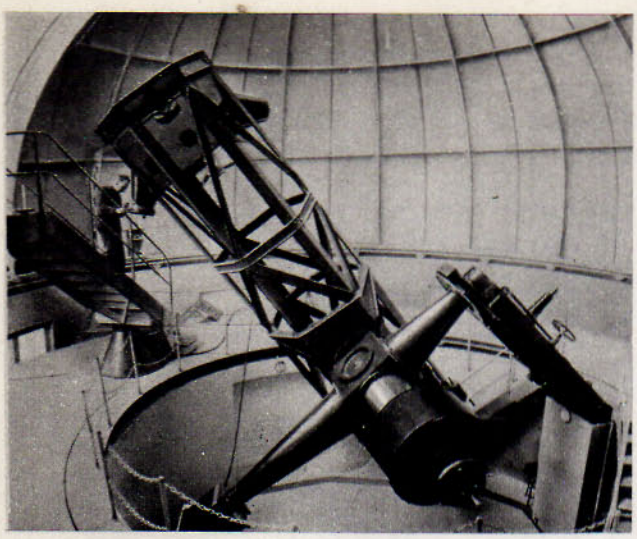


Fig. 744 — 47-INCH REFLECTOR OF THE HAUTE-PROVENCE OBSERVATORY. *(Photo, Laffineur)*

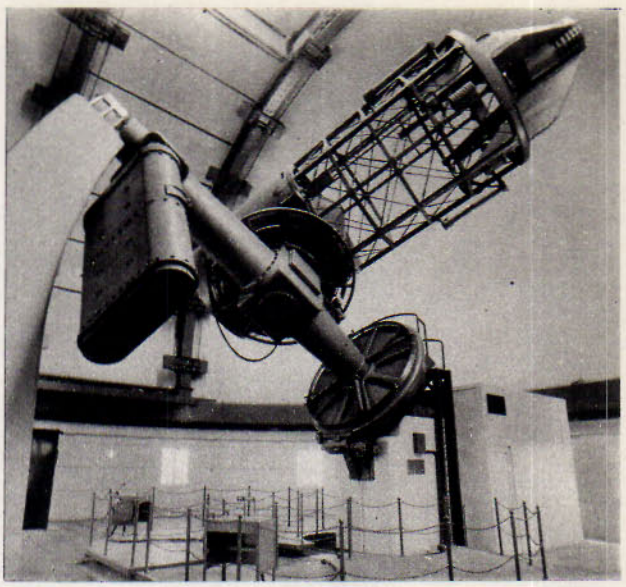


Fig. 745 — 82-INCH REFLECTOR AT THE McDONALD OBSERVATORY, FORT DAVIS, TEXAS. Note the unusual position of the counterpoise, and the curious supports of the photographic chassis at the prime focus.

For this reason compound object glasses, adequately corrected for the abaxial aberrations (coma, astigmatism and field curvature), were for a long time the only objectives that could be used in wide-field cameras — instruments that are invaluable for recording extended objects such as the Milky Way or comets.

This situation was revolutionised in 1930 with the invention of a new optical system by the German astronomer B. Schmidt, of Hamburg Observatory, which provides good definition over a very wide field; this it achieves by the combination of a spherical mirror and an aspherical correcting plate whose aberrations are equal but opposite to those of the mirror (Fig. 745). Numerous Schmidt telescopes (or, as they are more usually called, Schmidt cameras, since they cannot normally be used for visual observation) have been constructed in all parts of the world during the past 25 years, and have greatly facilitated surveys of large expanses of sky and statistical investigations. The numerous discoveries of supernovae made at Mount Palomar since 1936 were the result of systematic work with their 18-inch Schmidt camera.

The list of the largest Schmidt cameras in the table below is unlikely to remain up-to-date for long, since large numbers of these instruments are being constructed all the time and in many countries:

A very important characteristic of these instruments, which distinguishes them from the types previously considered, is that the usable field *increases* in size as the relative aperture increases. Whereas in telescopes of conventional type the aberrations (coma and astigmatism) limit the permissible inclination of the rays to the axis, in the Schmidt system, which is completely free from these aberrations, the angular diameter of the usable field increases in the same proportion as that of the aperture to the focal length of the mirror. Thus Schmidt cameras working at $f/3$, $f/2$, $f/1$, or even less (i.e. the aperture is larger than the focal length) in the case of smaller models designed for spectrographic work, are perfectly practicable. The limit for parabolic reflectors, on the other hand, is normally about $f/5$ or $f/6$,* and for 'astrographic' doublets (astronomical camera lenses) is rarely better than $f/7$ or, exceptionally, $f/5$.

* The exceptionally large relative aperture of the Mount Palomar reflector (f 3.3 is made possible by the use of a Ross correcting lens interposed in the beam converging from the mirror, just in front of the photographic plate.

THE WORLD'S LARGEST SCHMIDT CAMERAS
(Aperture > 16 inches)

OBSERVATORY	APERTURE OF CORRECTING PLATE (inches)	DIAMETER OF MIRROR (inches)	FOCAL RATIO (f/-)	DATE OF COMPLETION
Mount Palomar (USA)	48	72	2.5	1948
St Andrews	38	—	—	—
Uccle (Belgium)	33	47	2.5	1958
Harvard (S. Africa)	32	36	3.8	1948
Hamburg (Germany)	31.5			1955
Tonanzintla (Mexico)	26	31	3.5	1942
Kivu (Republic of Congo)	25.3	38	3.0	under construction
Specola Vaticana (Italy)	25	38	3.76	1957
Warner and Swasey (USA)	24	36	3.5	1943
Harvard (USA)	24	33	3.5	1940
Lowell (USA)	24	31	3.5	under construction
Mount Palomar (USA)	18	26	2.0	1936
Cambridge (England)	17	24	3.76	1953
Edinburgh (Scotland)	16	24	3.75	1951

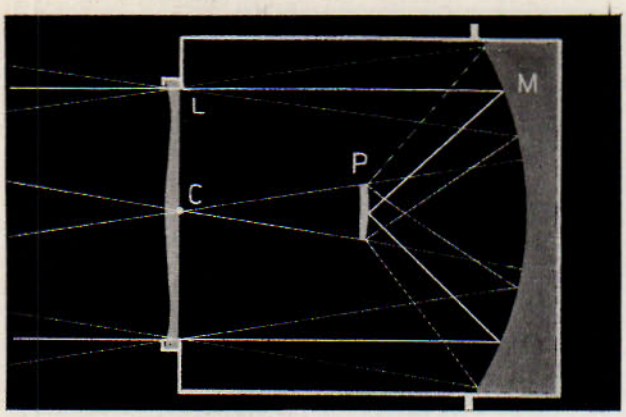


Fig. 746 — PRINCIPLE OF THE SCHMIDT CAMERA. The spherical aberration of the wide-aperture spherical concave mirror M is corrected by the aspherical plate L situated at its centre of curvature; a sharp image of a very wide field is formed on the spherical surface of the plate-holder P in the focal plane of M.

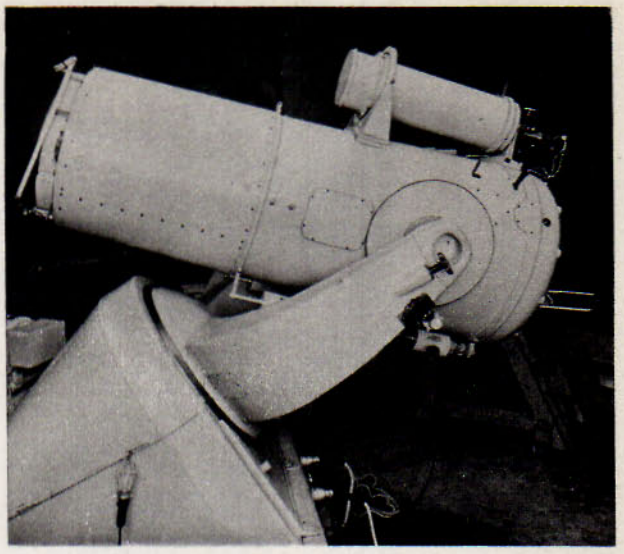


Fig. 747 — THE 18-INCH SCHMIDT CAMERA AT MOUNT PALOMAR, WITH WHICH A LARGE NUMBER OF SUPERNOVAE HAVE BEEN DISCOVERED

Now the photographic 'speed' of an instrument, which determines the minimum exposure required to record faint extended objects such as nebulae and comets, is primarily determined by its relative aperture; this fact explains the enormous interest that has been shown in these new instruments, for they can often secure in a few minutes images that previously had demanded exposures of several hours.

One minor disadvantage of the original Schmidt system however, is that the focal surface, in which the image is formed, is not a plane but spherical. This normally prohibits the use of glass photographic plates, which have to be replaced by film which is forced to the requisite curvature by pressure against a suitable support. Even this drawback has been elimated in some of the many optical combinations that have been developed from the prototype Schmidt during recent years.

With the exception of a few instruments with specialised functions, which we shall describe later, all the refracting and reflecting telescopes in our observatories are supported on mountings of the so-called equatorial type. There are many variants of this type of mounting, but they all have the following features in common: a 'polar axis' on the celestial pole (therefore parallel to the Earth's axis of rotation), and a 'declination axis' which is perpendicular to the polar axis. The telescope is mounted on the declination axis and can therefore be pointed to any part of the sky merely by rotating the two axes. Rotation of the declination axis will give it an elevation corresponding to the declination of the required object, while rotation of the polar axis will permit it to be directed towards the great circle passing through the object, which is defined by the angle (the so-called Hour Angle) that it makes with the observatory's meridian; this angle is determined by the object's R. A. and the time at which the observation is made. By means of graduated circles, clamped to the two axes, the telescope can thus be quickly directed towards any object (even though it may be invisible to the naked eye) merely by setting its co-ordinates on the circles.

The larger instruments are in addition usually fitted with one or more smaller telescopes, known as 'finders', by means of which the object can be centred in the much smaller field of the main instrument; they are mounted on the tube of the big telescope and adjusted so as to be accurately parallel to its optical axis. Other small viewing telescopes are mounted in various positions within convenient reach of the observer (usually, like the finders, on the tube of the main instrument) to enable him to read the graduations of the distant Hour and Declination circles.

Finally, when the instrument has been pointed towards the required object and clamped in this position, it must still be kept trained on the object, for the latter's position in the sky is being constantly altered by the Earth's rotation. This is achieved by means of the telescope drive, which is transmitted to the instrument via a large toothed wheel mounted on the polar (or hour) axis; the slow and uniform rotation of this axis is produced by means of a tangential worm, or endless screw. This in turn is driven by a small electric motor whose speed is commonly regulated by a very accurate pendulum clock which controls the motor's electrical input; alternatively, as in some of the large instruments installed in recent years, control of the speed of the drive is effected through a more complex arrangement comprising a synchronous motor, the frequency of whose input is regulated by a standard vibrator (quartz crystal).

Despite the drive, however, various factors usually conspire to make the rotation of the instrument less perfectly uniform than that of the Earth, with the result that the object will appear to drift slowly and more or less regularly across the telescopic field. These disturbing factors include a variety of mechanical imperfections in the mounting (flexure, etc.) or the drive (eccentricity of the axes, gears, etc.). But even if the instrument and its mounting were mechanically perfect the effect of atmospheric refraction — which varies with the object's altitude and also with such meteorological factors as temperature, pressure and humidity (see Chapter 4) — will displace the object from its theoretical position. For all these reasons telescopes are provided with electrically or mechanically operated 'slow motions' which enable the observer to correct any minor wanderings of the image and maintain its position in the field with the required precision: this is very great in the case of long-exposure photography, when a displacement of the images amounting to only a few thousandths of a millimetre is enough to produce appreciable distortion.

The observer is usually supported in a comfortable position at the eyepiece on some sort of movable platform or bridge; occasionally the height of the entire observatory floor is adjustable. In the case of the Mount Palomar

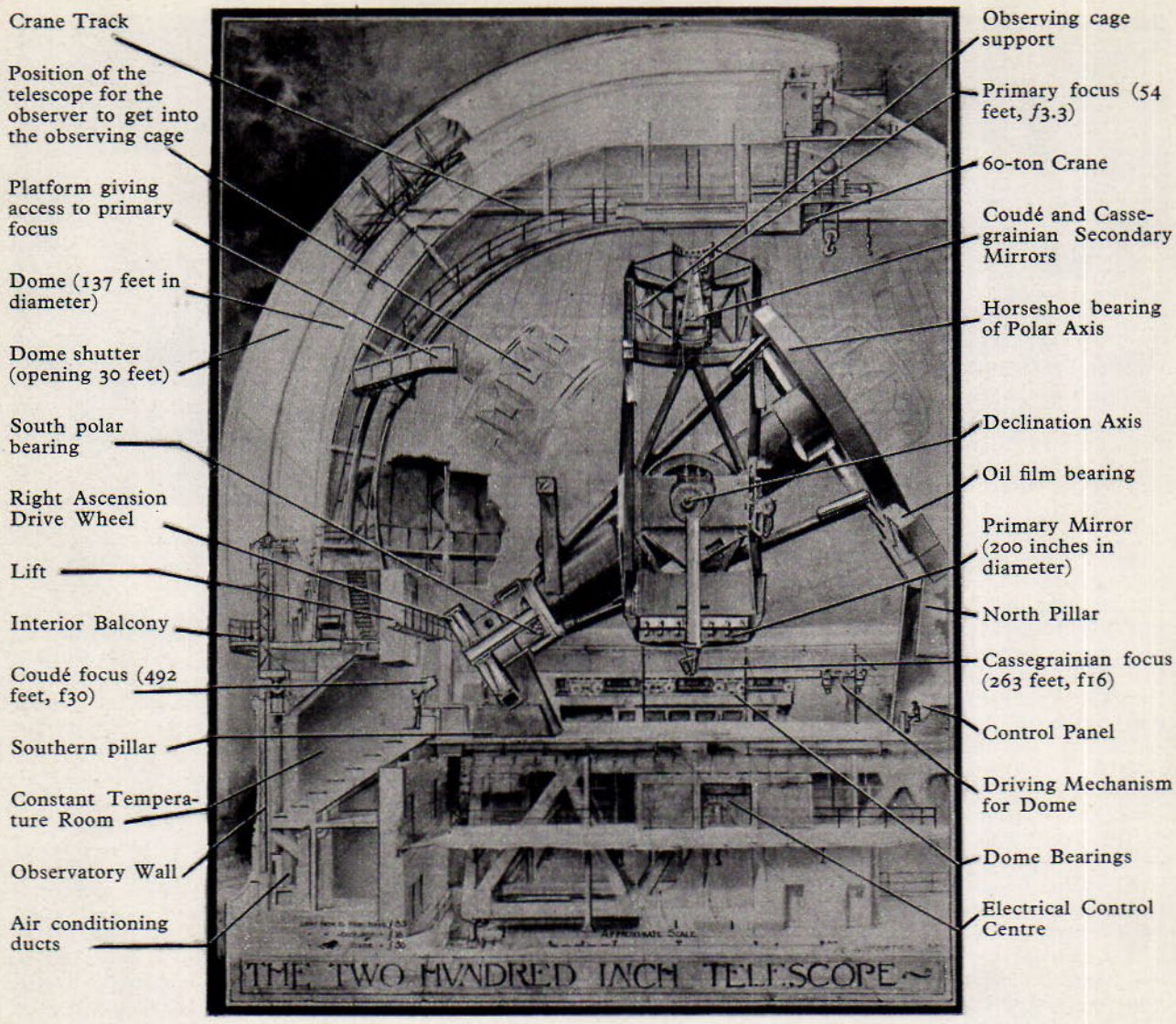


Fig. 748 — SECTIONAL VIEW SHOWING THE INTERNAL ARRANGEMENT OF THE DOME OF THE GIANT TELESCOPE AT MOUNT PALOMAR. *(Photo, California Institute of Technology)*

200-inch the aperture is so great that the observer can be accommodated in a small cage situated actually inside the telescope tube itself, just above the prime focus.

The whole instrument — telescope tube and mounting — is normally housed inside a hemispherical dome pierced by a slit which is covered by laterally sliding shutters when not in use; this slit uncovers a strip of sky which is wider than the rotating aperture of the instrument and stretches from the horizon to the zenith. By rotating the dome as a whole (this is operated electrically and is controllable by the observer), any part of the sky can be made accessible. The domes covering large instruments may be of very considerable size and weight: the diameter and height of the 200-inch dome, for instance, are 137 feet, and its weight 1,000 tons.

A fully equipped observatory will also have such auxiliary establishments as photographic and physics laboratories, a library, workshops etc., not to mention the living and administrative quarters necessary for the life and work of the resident and visiting astronomers.

European observatories are for the most part under the direction of either the State or some University; but a fair number, particularly in America, are supported by funds contributed by private organisations or foundations.

Positional measurements in astronomy. — The basic astronomical instruments — the refractor and the reflector — have been modified in numerous ways to suit the particular types of observation that are required of them.

Measurements of celestial positions and motions are made almost exclusively with refractors rather than with reflectors, on account of the former's comparative stability and insensitiveness to temperature changes. The fundamental instrument in classical positional astronomy — to which celestial mechanics owes its well-merited reputation for precision — is the meridian circle, whose functions is to measure the absolute positions of stars upon the star sphere. It is a refractor, usually about 10 inches in aperture, whose horizontal axis is supported by two massive piers bedded deep in the ground; between these it is free to swing, with the greatest precision, in an exactly vertical plane whose north-south orientation is also adjusted to the limits of achievable accuracy. At the eye-end of the telescope is fitted a micrometer (see p. 459) consisting of a series of vertical webs, one of which is

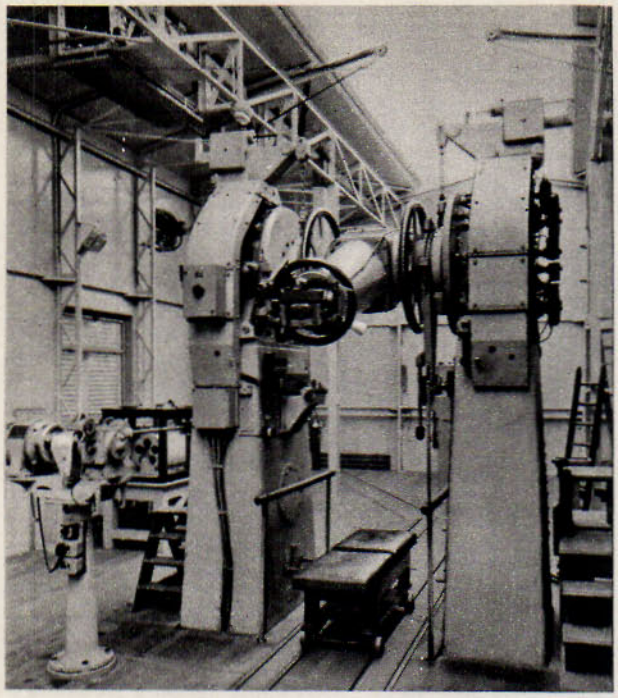


Fig. 749 — THE MERIDIAN TRANSIT CIRCLE AT THE ROYAL GREENWICH OBSERVATORY, HERSTMONCEUX. *(Photo, Royal Greenwich Observatory)*

movable, and two horizontal webs, one of which again is movable. The horizontal axis about which the instrument can rotate is fitted with one or two large-diameter circles, graduated with the highest precision possible. The small but inevitable residual errors of these circles are themselves determined once and for all by a long and systematic series of preliminary observations.

The installation is completed by a clock giving the local sidereal time which, it will be remembered, is the time as determined by the Earth's true period of rotation relative to the stars.

We have not space to describe the technical procedure followed by the meridian observer in carrying out his routine and extremely precise work, but we must emphasise its fundamental importance, and the fact that such observations are still and continuously being undertaken: the accurate determination of the absolute positions of the stars — often dull and unspectacular work — is the essential foundation on which astronomy is built. Though hardly likely to lead to new and exciting discoveries it remains today, as it has always been, the unwavering preoccupation of astronomers who seek an ever more precise definition of a system of co-ordinates which, amid the dismaying complexity of the cosmic maelstrom, may be accepted as relatively unchanging.

Absolute measurements of celestial co-ordinates are, like all absolute measurements, less accurate than differential measurements made between neighbouring objects — for example, the components of a double star. Measurements of this type are made with that venerable instrument the filar micrometer, mounted in the focal plane of an equatorial refractor. This micrometer consists of a flat, rectangular box containing, in theory, two (but often several) taut spider webs at right angles to one another, and one or two others, parallel to one of the former, which are mounted in a movable frame. The observer can adjust the position of this frame as he wants by turning a knob connected with a screw and carrying a graduated drum from which he can read the number of turns and fraction of a turn that have been applied to the screw. The box is mounted on a circle graduated in degrees by means of which its orientation, and that of its webs, can be read off. At night the webs are rendered visible against the dark background of the sky by suitable side lighting.

Again without going into details of the procedures followed in making these measurements, it must nevertheless be pointed out that their accumulation has made possible the computation of the orbits of binary stars, and of the masses of their components; on such measurements, in the last analysis, rests a large part of stellar physics and dynamics.

These measurements are in fact much more error-prone and difficult to make than the foregoing simplified description might suggest. Owing to the incessant agitation of the atmosphere (and, possibly, small irregularities in the telescope drive) a star will appear to dance continuously back and forth across the web which the observer is trying to adjust precisely over its image. This is so difficult a feat that it may be likened to the task of keeping a rifle accurately trained on one of those ping-pong balls bouncing on a jet of water that one sees in fairground shooting galleries. And, just as not everyone is a natural first-class shot, so not every observer is capable of making accurate double-star measurements.

The accuracy of the measurements and the value of the conclusions to be drawn from them are for this reason considerably impaired, and attempts have been made to replace this visual method by a photographic one. When the photographic images of the two stars are fairly well spaced, and are neither too bright nor too faint, the measurements are easy to make photographically. The developed negative is inserted in the carriage of a plate-measuring machine, which is adjustable in two mutually perpendicular directions by means of micrometer screws with graduated drums; the stellar images whose relative positions are to be determined are brought in succession under the web of the view microscope, and the difference of the readings of the micrometer drums at once gives the required quantities; the plate-scale has of course to be determined first, and this is done by making analogous measurements of the images of other stars whose positions have been independently determined by meridian observations. It is always necessary to apply small corrections for faults in the micrometer screws, variation of plate-scale from the centre to the edge of the negative, etc. Measurements of this kind, accurate to one thousandth of a millimetre, have provided the data tabulated in the *Astrographic Catalogue* based on the Carte du Ciel plates.

Telescopes that are to be really useful for photographic double-star measurements must have long focal lengths (30 feet or more) so as to provide an adequate focal plane scale; given this, as well as certain technical refinements which need not be described, the photographic method yields a much higher standard of accuracy than the visual — 0.01″ as against 0.1″. It is this high degree of precision that has, for example, revealed the existence of the tiny perturbations caused by the invisible 'planetary' companions of 61 Cygni and 70 Ophiuchi (Chapter 13).

Unfortunately, the photographic method is not applicable to very close doubles, for the images of the components formed by the light-sensitive emulsion become confused with one another; even when the separation is sufficient for the individual images to be clearly distinguished visually, though almost in contact, certain photographic and optical effects tend to falsify the measurements (Kostinsky effect). To deal with such stars, special visual micrometers, more accurate than the ordinary filar micrometer, have been developed during the

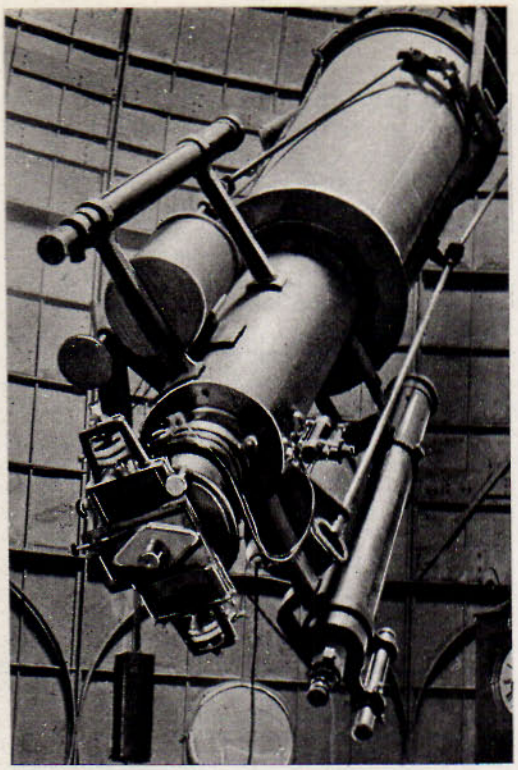


Fig. 750 — DOUBLE STAR MICROMETER ATTACHED TO THE 15-INCH REFRACTOR OF THE PARIS OBSERVATORY. *(Photo, L. Rudaux)*

past 25 years; the most important of these are Hargreaves's comparison-image micrometer, Muller's birefringent (double-image) micrometer, and Danjon's half-wave interferometric micrometer; these, however, can only be mentioned in passing, since their employment is by no means general.

We come, finally, to those doubles whose separation is so small that they cannot be measured with any of these instruments: the telescope is incapable of separating their images distinctly, so that they appear as a single, perhaps somewhat elongated stellar image. Recourse must then be had to the interferometric method. Its principle is as follows. If the objective of a telescope is covered by a diaphragm in which two holes or slits are cut at equal distances from its centre, it will be found that the diffraction images of the stars are crossed by a series of symmetrically arranged, alternately bright and dark fringes; these are produced as a result of interference between the light rays that have passed through the two openings (Fig. 751). This is Young's classical demonstration, which the reader can easily repeat for himself by looking at the filament of an electric bulb through two needle-holes pierced very close together in a piece of card. The greater the separation of the holes, the more numerous and closely packed are the fringes. In the case of a single star the fringes remain visible whatever the separation and orientation of the holes may be; but if the star is a double, each component gives rise to its own fringe system, which is displaced relatively to that of the other; and for a particular separation and orientation of the holes the bright fringes of one system will coincide exactly with the dark ones of the other, and vice versa. The image will then appear to be normal, no fringes being visible. If the separation of the holes is then increased or decreased, or if their orientation relative to the line joining the components is altered, the fringes will reappear.

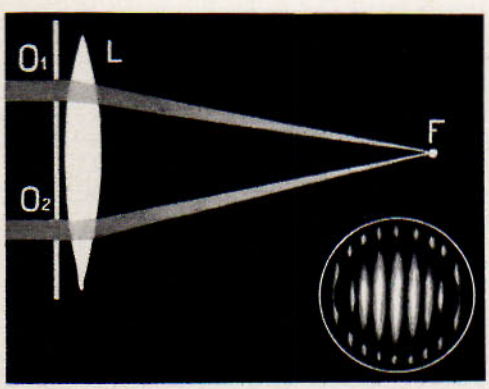


Fig. 751 — PRINCIPLE OF INTERFEROMETER MEASUREMENTS, AND THE FORMATION OF YOUNG'S FRINGES. The figure in the circle shows the appearance of the fringes superimposed on the diffraction pattern of a star (spurious disc and first ring) formed at the focus F of the lens L by interference between the two beams passing through the slits O_1 and O_2. If the star is double, or if it produces a sensible disc, the fringes produced by the two components, or by different points on its surface, overlap and produce mutual interference.

It will be evident, without going into details, that a micrometer could be designated to exploit this principle, and that it could be used to measure doubles of very small angular separation. But in the form just described, the distance separating the holes is limited by the aperture of the telescope's objective. To overcome this restriction, and thereby greatly increase the scope of the method, the American physicist Michelson fixed a rigid metal girder, 18 feet long, across the mouth of the Mount Wilson 100-inch telescope, and on it he mounted four plane mirrors, symmetrically disposed on either side of the optical axis and inclined to it at an angle of 45°. The two outer mirrors reflected the light from the star to the inner pair, which in turn reflected the two beams down the tube to the 100-inch objective (Fig. 752). The outer mirrors, which thus take the place of Young's two apertures, could be moved apart from one another to the maximum distance of 18 feet.

It was with the help of this instrument that F. G. Pease

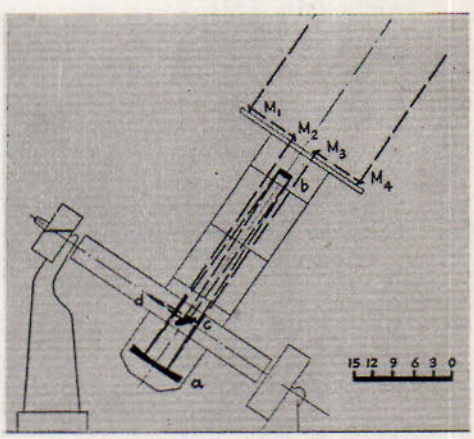


Fig. 752 — DIAGRAM OF MICHELSON'S INTERFEROMETER, MOUNTED ON THE MOUNT WILSON 100-INCH REFLECTOR. The mirrors M_1, M_2, M_3, M_4 transmit the two beams which produce the mutual interference at *d* (the Coudé focus of the telescope) after reflection at the mirrors *a*, *b*, *c*.

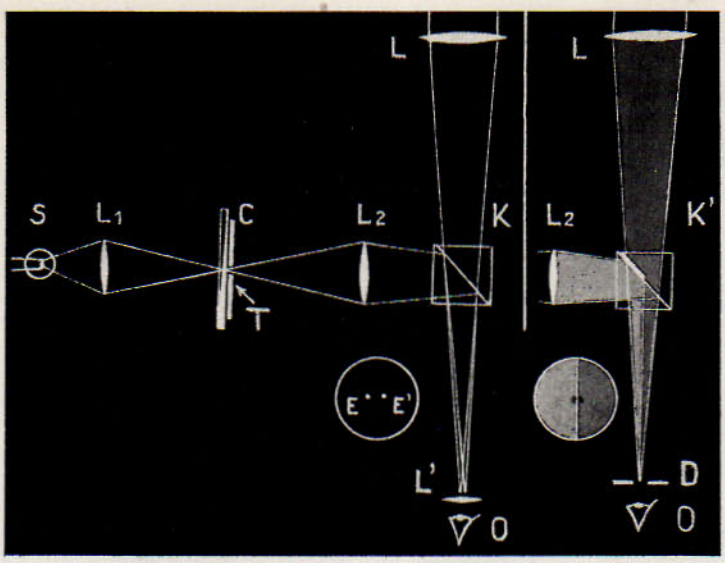


Fig. 753 — DIAGRAM ILLUSTRATING THE PRINCIPLE OF THE VISUAL ASTRONOMICAL PHOTOMETER: left, employing an artificial star; right, illuminated exit pupil. The lens L_1 forms a reduced image of the electric filament S on a pinhole T, in front of which is placed the optical wedge C. The image of the illuminated pinhole is projected by lens L_2 into the field of the ocular L′, after reflection at the half-silvered diagonal of the optical cube K, and constitutes the artificial star E′; its brightness can be varied by adjusting C, to match that of the image of the real star E, formed alongside E′ by the telescope objective L.

On the right is shown the simple way in which it can be transformed into a photometer without a diffusing screen, by replacing the half-silvered cube K by a second cube K′, only half of whose diagonal is silvered, and the ocular L′ by a simple lens-cap D, behind which the eye is placed. The appearance of the field is shown in each case.

in 1920 succeeded in measuring not only the separations of binaries such as Mizar and Capella (0.05″ or less — too small to be visually perceptible with any telescope in existence) but also the angular diameters of certain supergiant stars. Different sets of interference fringes are in effect produced by different points on the disc of such a star, and the superposition of these systems will lead to the disappearance of the fringes if the outer mirrors are far enough apart. Even so, these very delicate observations were successful only in the case of half a dozen suitable stars (see Chapter 14).

Before leaving the subject of astronomical measurements of position, something must be said of two other very important applications: the measurement of the distance of the Sun from the Earth, and of the distance of the stars.

The Sun's distance is calculated from the measurement of its parallax, the principle of which has been described in Chapter 12. The base of the triangle to which the measures are referred — the Earth's equatorial radius — has been determined with great accuracy in terms of the standard metre by geodetic surveys of the Earth's surface.

Since it is not possible to pinpoint the Sun with sufficient accuracy for the direct measurement of its parallactic displacement (which is always less than 10″), owing to its angular size and the ill-defined nature of its limb, it has long been the practice to employ one of the nearer planets as stepping-stones: the small images of Mars and Venus facilitate the accurate determination of their position, and their parallaxes are relatively large — 20″ to 30″ when they are nearest the Earth. The relative dimensions of the various planetary orbits are known with great accuracy from Kepler's Third law (p. 16), and it is therefore necessary to determine only the linear distance of any one of them from the Earth in order to establish the scale of the whole Solar System — and, in particular, the mean linear distance of the Earth from the Sun. This quantity, it will be remembered, serves as the Astronomical Unit of distance.

But, unfortunately, when either Mars or Venus is

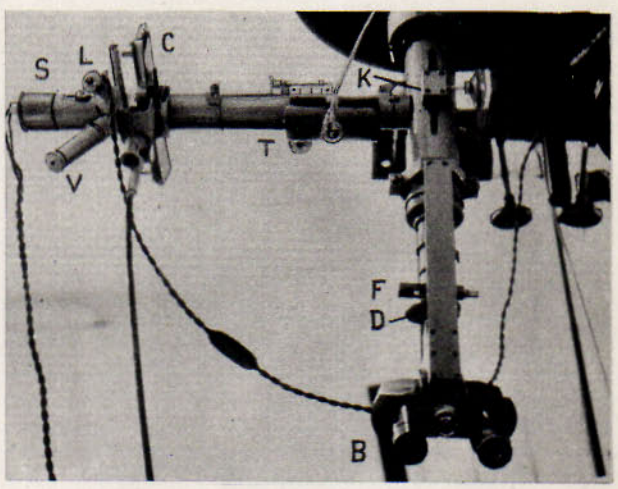


Fig. 754 — VISUAL PHOTOMETER MOUNTED ON THE 8-INCH REFRACTOR OF THE HOUGA OBSERVATORY. Compare with the diagrammatic layout of Fig. 753. The tube T contains the lens L_2 and an auxiliary optical unit forming the artificial star; F is a coloured filter holder, and B a binocular eyepiece which facilitates the observation; V is a reading microscope focused on the graduations of the optical wedge.

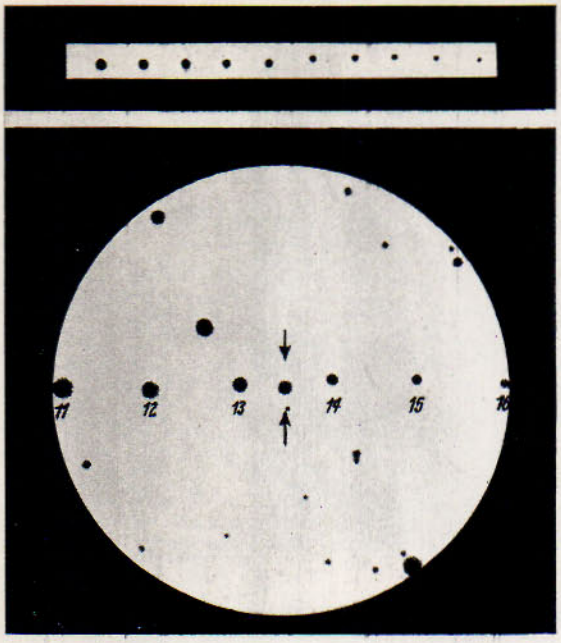


Fig. 755 — PHOTOGRAPHIC STELLAR PHOTOMETRY BY MEANS OF A GRADED SCALE OF IMAGES. Above: appearance of the comparison scale; below: appearance of the field, showing the graduated scale superimposed on a small region of a negative bearing the image of the star to be measured.

nearest to the Earth (the former at perihelic opposition and the latter during solar transit) their large angular diameters, 25″ and 65″ respectively, combine with various ancillary factors to make accurate measurements of position difficult and their reduction uncertain; such ancillary factors include the red colour of Mars and the brilliance of its polar cap, and in the case of Venus in transit its atmospheric halo, the optical phenomenon known as the 'black drop', and the 'boiling' of the image of the Sun's limb due to atmospheric turbulence. For these reasons the results obtained up to the end of the 19th century were very discordant, as can be seen from the following table, which cites some of the values obtained for the Sun's parallax:

SUCCESSIVE VALUES OF THE SOLAR PARALLAX

DATE	PARALLAX (″)
c. 1700	10 ±
c. 1800	8.7 ±
1825—55	8.57
1855—85	8.86
Since 1890	8.80

But about 60 years ago the English astronomer D. Gill hit on the happy idea of using some of the minor planets which, since they always appear as stellar points, lend themselves to much more precise measurement than is possible in the case of planetary discs. This method owes its success very largely to the discovery in 1898 of the asteroid Eros (433), which at times approaches closer to the Earth than any of the planets; its parallax may, in fact, be as large as 60″. Two international programmes have been organised for its observation: the first, in 1900—1 (maximum parallax, 28″), and the second in 1930—1 (maximum parallax, 50″). They provided the results shown in the next column:

DATE OF AUTHORITY	SOLAR PARALLAX (″)
1900—1 A. R. Hinks	8.806 ± 0.004 (p. e.)
1930—1 H. Spencer Jones	8.790 ± 0.001 (p. e.)

The latter value is based upon 2,847 photographs, taken by 30 different instruments at 24 observatories in 14 countries, and gives the mean Earth-Sun distance as

93,005,00 ± 9000 miles.

The most recent re-determination of this parallax is due to E. Rabe, who in 1950 established a value of 8.7984 ± 0.0004 (p. e.). The Sun's parallax can also be evaluated by physical and dynamical methods, about which nothing further need be said here (see Chapter 20): they give a mean value of 8.803 ± 0.001″; the small discrepancy between this figure and Rabe's parallax value is at present unexplained.

The determinations of stellar parallax are much less precise than that of the Sun's parallax because of the smallness of the angles involved, which in no case exceed 1″. The trigonometrical method is in fact only possible for a few of the nearest stars, as we have already seen. Their positions are determined micrometrically — originally by direct measurements of absolute positions, nowadays by photography, relative to nearby but more distant stars. The relative parallax so obtained is then corrected for the very small probable mean parallax of the background stars.

We have seen how other methods of distance determination have enabled us, little by little, to extend the surveyed domain to the limits of the visible Universe; but in the final analysis all these determinations depend on the few hundred directly measured trigonometrical parallaxes, and on the fundamental standard astronomical distance of the Earth from the Sun. It is thus easy to see why astronomers have expended so much effort on the accurate determination of the value of the Astronomical Unit.

Photometric measurements in astronomy. — The determination of the brightness of celestial objects in no less important than that of their positions.

We have already described briefly how the subjective impressions of stellar brightness have been systematised and made progressively more accurate, culminating in the modern scale of apparent magnitudes (Chapter 12)*. This final development only became possible after the invention of the first astronomical photometers, about a hundred years ago.

One of the most familiar and widely used is Zöllner's artificial star photometer, which is fitted to the eye-end of the telescope. It consists, in essentials, of a small lateral tube containing an electric bulb, a screen pierced by a fine hole which serves as the artificial star, a diverging lens which reduces the latter's image still further and two or three polarising elements (nicol prisms). The brightness of the artificial star can be varied at will simply by rotating one of these prisms about the optical axis of the system — the relation between brightness and angle being specified by Malus's law. Nowadays, in place of the nicol prisms a wedge of tinted glass is often employed; this transmits a fraction of the light from the artificial star, according to its position, and this is determined by previous calibration. A lens, followed by a plane-parallel

* The apparent magnitude (m) of a star is defined, in terms of the illumination (E) received at the Earth's surface, by Pogson's formula (1857):

$$m = -2.5 \log E - k$$

In the case of visual magnitudes, if E is expressed in lumens, the value of the constant k is —14.3; this is the apparent magnitude of a standard candle viewed from a distance of 1 metre.

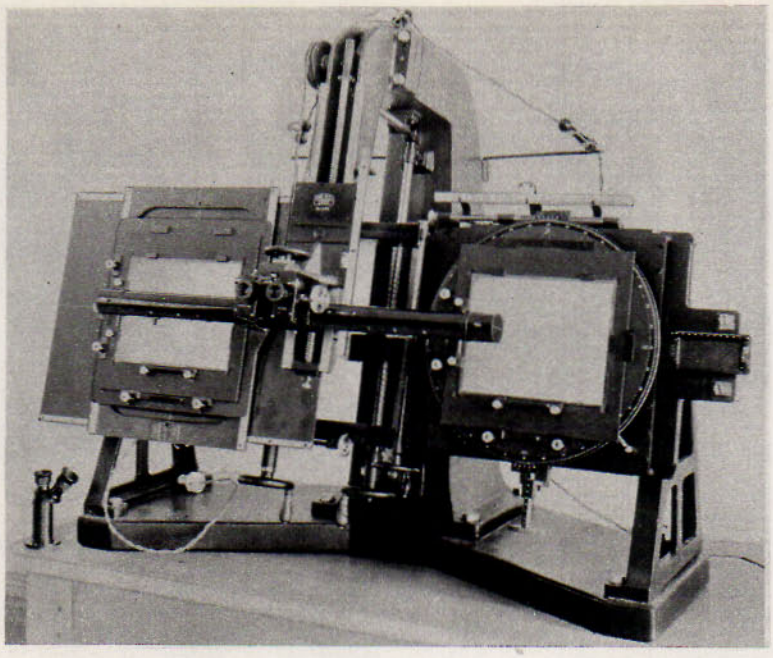


Fig. 756 — ZEISS BLINK MICROSCOPE AT YERKES OBSERVATORY

plate set at an angle of 45° to the tube, projects the image of the artificial star into the field of the ocular at the side of the image of the real star whose brightness is to be measured (Fig. 753).

It was with a Zöllner photometer that Müller and Kempf, at Potsdam Obervatory, between 1888 and 1905 collected the data for their great photometric catalogue of bright stars in the northern hemisphere — to mention only one of the many applications that have been made of this instrument.

A disadvantage frequently encountered with the Zöllner photometer tends to be rather troublesome: the artificial star has not exactly the same appearance as the real star, and this dissimilarity affects the accuracy of the measurement. This drawback is avoided in those photometers which, by means of mirrors or prisms mounted in front of the object glass, reflect the image of a real star into the telescopic field alongside that of the star whose brightness is to be measured. The best known instrument of this type is the meridian photometer, first constructed by E. C. Pickering in about 1880 at Harvard College Observatory; it derives its name from the fact that it permits the comparison of stars near the time of their meridian passage. In this instrument the light of each star is split into two beams by a birefringent prism; each star thus has two images (known as the ordinary and extraordinary images) whose light is polarised at right angles to one another; the brightness of these images can therefore be varied by the rotation of a nicol prims (the analyser) mounted in front of the ocular, just as in the Zöllner photometer. The observation consists in equalising the ordinary image for one star and the extraordinary image of the other, their magnitude difference being derived as before. More than a million individual observations, made with a series of such instruments constructed between 1880 and 1910, were the basis of the great Harvard catalogues, which tabulate the magnitudes of some 50,000 northern and southern stars to magnitude 10.

Of the numerous photometers that have been specially designed for the observation of variables, we shall mention only one: Danjon's 'cat's-eye' photometer. With this the image of the variable is viewed direct while that of the comparison star is projected into the field by means of two totally reflecting prisms, between which is a diaphragm with a square aperture (the cat's-eye); the size of this aperture is adjustable, thus allowing the quantity of light entering the field to be varied at will: the required magnitude difference is calculated from the length of the diagonal of the cat's-eye (read from a scale) when the two

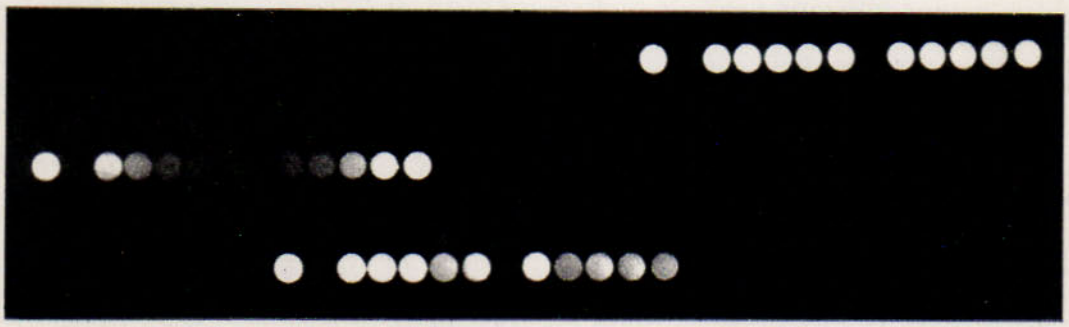


Fig. 757 — OBSERVATION OF A VARIABLE STAR BY MEANS OF EXTRA-FOCAL PHOTOGRAPHY. The brightness variation of the eclipsing binary RZ Cassiopeiae during the course of its eclipse (principal minimum) is clearly revealed by the comparison of its successive images with those of nearby non-variables. *(Photo, J. A. Parkhurst, Yerkes Observatory; 3 August 1907)*

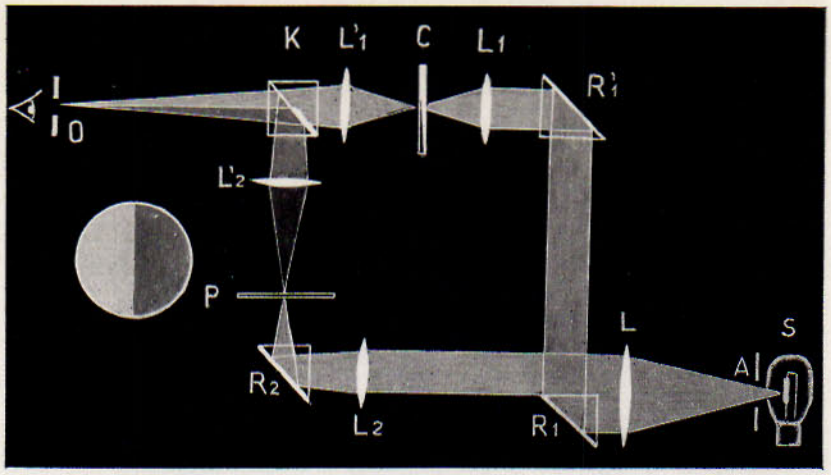


Fig. 758 — OPTICAL LAYOUT OF THE FABRY AND BUISSON VISUAL MISCOPHOTOMETER. Light from the mercury vapour lamp S passes through the aperture A and is focused into a parallel beam by the lens L. It is then divided by the totally reflecting prism R_1. One of the beams is focused by the lens L_2 on the photographic plate P, via the prism R_2; the other is directed by the prism R'_1 and the lens L_1 to the optical wedge C. The two beams, collected respectively by lenses L'_2 and L'_1 are finally directed to the observer's yey at O — and the other after reflection at the half-silvered surface of the optical cube K, **and the other** directly. The appearance of the field seen at O is shown in the circular inset. The measurement consists in equalising the two halves of the field by means of the adjustable wedge C.

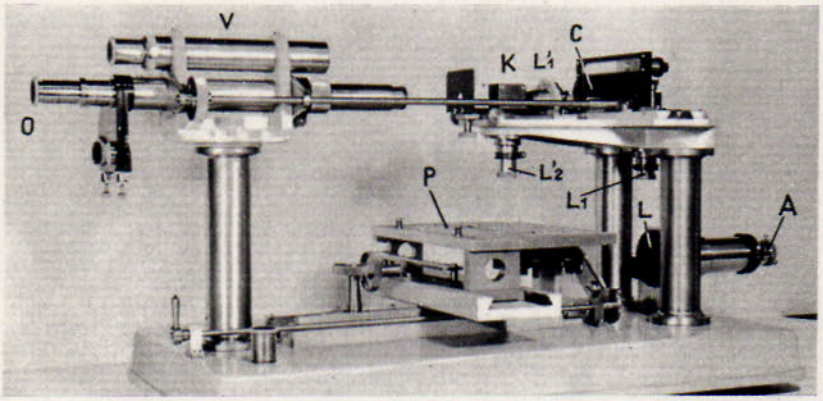


Fig. 759 — MODERN FORM OF THE FABRY AND BUISSON VISUAL MICROPHOTOMETER. Compare with the preceding diagrammatic layout. Behind the peephole O there is now a low-power telescope to facilitate the making of the observation; the viewer V is directed at the graduated scale of the optical wedge C. *(Photo, Baillet)*

images are equalised. This instrument not only gives very accurate results on variables, but has the added advantage of being able to measure the brightness of extended images — the Moon, for example, or the 'Earth light'; from systematic measurements of the latter Danjon was able to estimate the brightness of the Earth, and even its seasonal variations, as seen from outer space.

For measuring the brightness of extended images (particularly if these are faint, like those of the Milky Way or the night sky) a different photometric principle is used. The simplest of such photometers consists simply of a tube, like a telescope tube without object glass or ocular, fitted at one end with a centrally pierced cap to define the position of the observer's eye; the other, open, end is pointed at the sky or whatever the object may be. The periphery of the field of vision through the tube (or a band of some diffusing material stretched diametrically across it is illuminated by a small electric bulb, and the intensity of this illumination regulated by an optical wedge. The observation consists in adjusting the wedge until the comparison strip or ring is indistinguishable from the background that is being measured. Usually it is preferable to retain the object glass of the telescope, so as to define more precisely the area of sky that is being observed (its image being formed at the eye aperture) and to eliminate stars from the field. With the same optical set-up the brightness of stars can also be measured, and with improved accuracy; they will be seen as extended discs, the image of a sufficiently bright star, formed exactly at the eye-aperture by the object glass, appearing to flood the objective with uniform illumination (Fig. 753). This is the principle employed in Fabry's photometer which has no diffusing screen, of which more will be said later on.

Finally it should be mentioned that the magnitudes of long period variables, novae, etc., are generally determined by ordinary eye estimates, without the help of any photometric instrument. Such estimates are based on

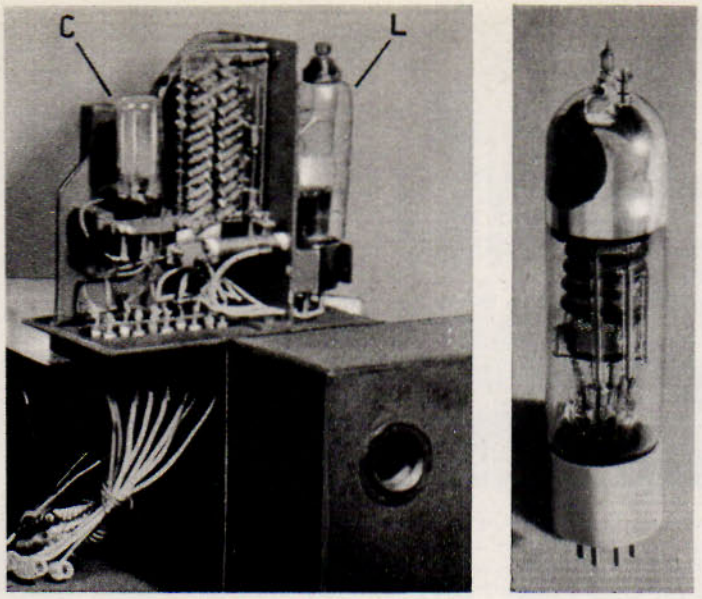


Fig. 760 — Left: photoelectric cell, C, and continuous current amplifier with special valve, L. *(Institut d'Astrophysique, Paris.)* Right: photoelectric cell incorporating an electron multiplier, constructed in the physics laboratory of the Paris Observatory. The series of amplifying stages can be seen, and, at the bottom, the collecting electrode. *(Photo, Lallemand)*

comparisons of the unknown star with nearby stars whose magnitudes have already been determined. The accuracy attainable in this way (0.1 to 0.2 magnitude) is perfectly adequate in view of the wide magnitude range of these variables.

But although visual photometry is still employed in certain cases, it has been increasingly replaced by photographic methods. The advantages of photography (in photometry as elsewhere) are the objectivity of impersonal observation, the permanence of the record — allowing its examination and measurement to be deferred to a more convenient time — the cumulative nature of the emulsion's reaction to the image falling on it, thus making stars accessible which are too faint to be observed visually, and, finally, the fact that large numbers of stars can be recorded simultaneously.

The earliest photographic method — and to this day the only one applicable to very faint stars — is that of focal plane photometry; it is based on the fact noted by G. P. Bond in 1857, in the very early days of astronomical photography, that the size of the image recorded by the emulsion is a function of the brightness of the star and the length of the exposure. The simplest way of utilising this fact, and the one that has given us the photometric data of the *Astrographic Catalogue*, it to place the negative to be measured on the stage of a low-power microscope into whose field is inserted a scale of graded star images that have been obtained with exposures increasing in geometrical ratio: if, for example, each image had been given 2.5 times the exposure of its neighbour, then they would differ one another by an interval of 1 magnitude. The measurement then consists simply of selecting two adjacent images on the scale, such that one is just larger and the other just smaller than the unknown image; the scale position of the unknown can be interpolated by eye with an accuracy of about 0.1 magnitude (Fig. 755). In this way the magnitude of every star on the plate can, in theory, be determined, provided that some are known independently so that the scale can be calibrated. In practice, as always, certain corrections (often complex) have to be applied, depending on the distance of the image from the centre of the plate, the colour of the star,

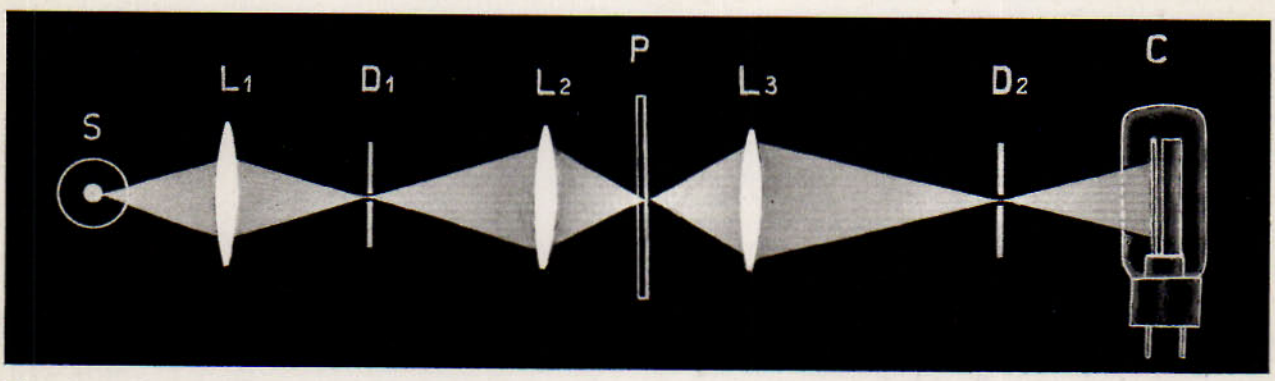


Fig. 761 — DIAGRAM ILLUSTRATING THE OPTICAL PRINCIPLE OF THE RECORDING PHOTOELECTRIC MICROPHOTOMETER. Light from the artificial source S, having passed through the photographic plate P, is projected on to the photoelectric cell C, whose amplified current produces a deviation of the galvanometer needle, which is recorded. The function of the diaphragms D_1 and D_2 is to limit the scanned region of the plate and to eliminate stray light.

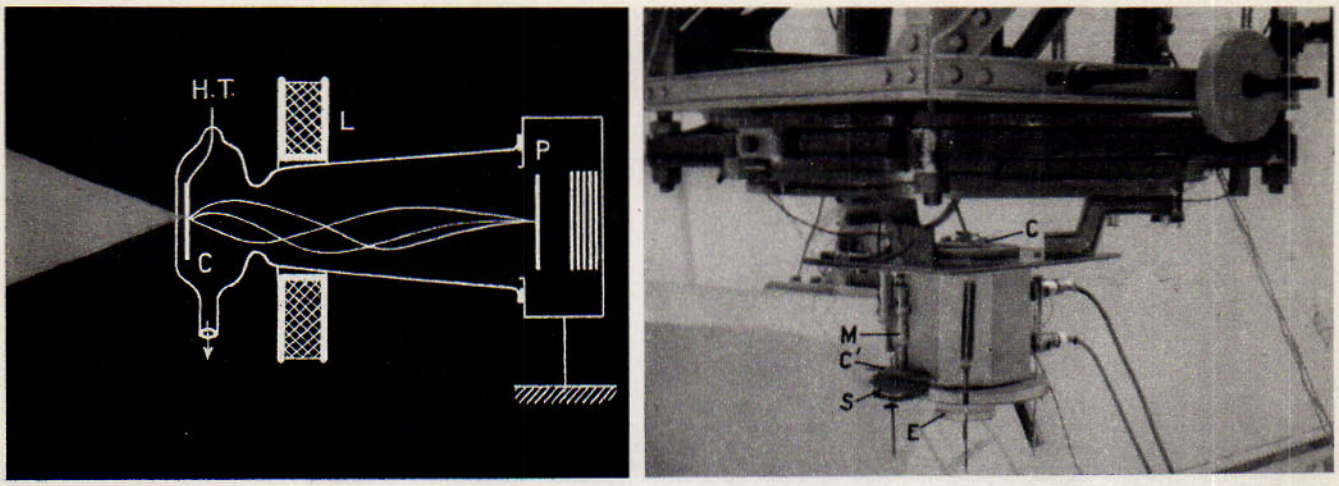


Fig. 762 — THE ELECTRONIC TELESCOPE. Left: principle of its operation. The translucent photoelectric cathode C, upon which the light collected by the visual telescope is focused, emits electrons which are accelerated in the field set up by HT, and deviated by the magnetic field of the coil L to a focus on the photographic plate P (or on a fluorescent screen in the case of visual observation). Right: electronic telescope mounted at the Cassegrain focus of the 32-inch telescope of the Haute-Provence Observatory. Compare with the diagrammatic layout: C, translucent photoelectric cathode; E, fluorescent observing screen; to the right, the HT and magnetic focusing field input leads; to the left, the auxiliary light source S and the microscope M, which provide the artificial comparison star whose brightness is adjusted by means of the optical wedge C′, for the study of variable stars. *(Photo, F. Lenouvel, 1947)*

Fig. 763 — SPECTRUM OBTAINED USING AN ORDINARY (KODAK 103C) PLATE, EXPOSURE 250S.

Fig. 764 — SPECTRUM OBTAINED WITH AN ELECTRONIC (ILFORD G5) PLATE, EXPOSURE 10S

and other factors.

This method, though undeniably rapid, is not one of the most accurate; the main sources of inaccuracy are irregularities in the distribution of the silver grains (as well as small local defects) in the light-sensitive emulsion, and the fact that the diameters of the images increase rather slowly as a function of magnitude.

Mention should be made at this point of the stereo-comparator or blink microscope — an instrument by means of which two plates of the same region, taken with the same instrument at different times, can be viewed alternately at the rate of a brisk flicker; this rapidly alternating presentation of the two plates to the field of vision greatly facilitates the search for variables, novae, etc., for if an image occurs on only one of the plates it will appear to 'blink' (i.e. alternately disappear and reappear) — an effect that catches the eye very easily, even in the midst of a plate crowded with stars. Pluto was discovered by the examination of plates with a blink microscope.

A more accurate photographic method than that of focal plane photometry was suggested by the French astronomer Janssen about 1880; it is, however, only applicable to comparatively bright stars. In this extrafocal method the plate is not exposed in the focal plane but slightly in front of or behind it; each star image is thus expanded into a circular disc whose diameter depends on the distance of the plate from the focal plane, and whose degree of darkening (on the negative) is a function of the star's brightness. Also recorded on the plate is a series of images produced with the aid of a calibrated optical wedge, or any other device for providing images of known relative brightness; the measurement then

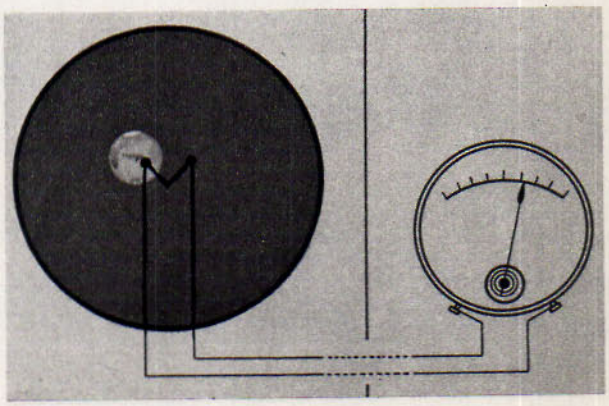


Fig. 765 — PRINCIPLE OF THE MEASUREMENT OF PLANETARY TEMPERATURE WITH A THERMOCOUPLE

consists merely of identifying among the scale images those with the same degree of darkening as that of the two stars under consideration, and from the known relation between their brightness the magnitude difference of the stars can be calculated.

This work of comparing photographic star images is facilitated, and its accuracy improved, by the use of a densitometer or microphotometer. In its usual form this instrument permits the density of the star images to be compared by means of a graduated optical wedge. Of the visual instruments of this type, that developed by the German astronomer Hartmann in 1899 has seen much service: in the field of its eyepiece are projected, side by side, the image of the optical wedge and of the area of the plate whose density is to be measured; these are illuminated by the two halves of a split beam from a small electric bulb. The wedge is then adjusted so that the illumination of the two halves of the field is equal. This arrangement has one serious disadvantage: the photographic grain can be seen in the magnified image of the negative, whereas the image of the wedge is uniform in texture, and this dissimilarity of appearance makes accurate comparison much more difficult. Different methods, of varying efficacy, have been devised to reduce this trouble, but it has only been wholly eliminated in the microphotometer developed by the French physicists Fabry and Buisson in about 1920, and widely employed since that date. Here the images of the optical wedge and of the negative are projected on a simple peep-hole, such as a lens cap, without any ocular being included; when the eye is placed behind this it will observe the two halves of the field to be uniformly illuminated, as explained earlier (p. 446). Under these circumstances their accurate equalisation is a much simpler matter (Fig. 758).

It should be noted, too, that this form of apparatus lends itself to the measurement of focal plane images, the field appearing uniform no matter how irregular the darkening of the plate interposed in the beam of light may be: the measured quantity is then the fraction of the section of this beam that is obstructed by the stellar image situated at its centre. Results obtained in this way are about twice as accurate as those obtained with a comparison scale — mean error of 0.05 as compared with about 0.1 magnitude.

Despite the fact that the extrafocal method gives an accuracy of $\pm$ 0.03 to $\pm$ 0.05 magnitudes, certain precautions being taken, it nevertheless has one drawback: the extrafocal sections of cones of rays emerging from the objective are rarely completely homogeneous, and these irregularities reduce the accuracy of the measurements.

To overcome this difficulty, Fabry in about 1910 proposed an alternative system, the advantages of which have secured its widespread adoption; it is a photographic application of the principle employed in the visual photometer without a diffusing screen. Just in front of the focal plane is mounted a screen, through a small hole in which the light from the selected star passes to a small, short-focus photographic objective located in the focal plane itself; this projects on to the plate or film a much reduced image of the telescopic objective filled with the light of the star; this so-called pupillary disc is perfectly uniform, and allows determinations to be made with an accuracy of $\pm$ 0.01 or $\pm$ 0.02 magnitude. An additional merit of this technique is that it is equally applicable to extended objects such as nebulae, comets, star clusters and even the faint luminosity of the night sky.

Of the many accessories employed in photometry, only one can be described here — the photometric grid, which is used to imprint on the photographic plate a series of images of known magnitude difference. It consists of a grid of thin, parallel, equidistant metal wires mounted in a frame and placed in front of the photographic objective. It has the double effect of reducing (by an amount that can be calculated or measured in the laboratory) the quantity of the light transmitted, and of producing low-dispersion diffraction spectra alongside the image of each star (see Chapter 20); the first two of these spectra, in particular, are practically indistinguishable from the central image, though fainter than it by an amount that likewise is calculable or measurable. This apparatus, first suggested by the Danish astronomer Hertzsprung in about 1905, has been extensively used since that time.

Electrical measurements in astronomy. — The general use of photography in astronomy, dating from about 1880 to 1890, has brought about a revolution in the observational field, and has in large part been responsible for the majority of recent discoveries.

It seems likely that the application of the varied resources and techniques of electricity — or, rather, electronics — whose use has become increasingly widespread in astronomy during the past thirty years, will lead to at least equally spectacular progress. Its consequences are still difficult to foresee, but their scope will certainly be immense. The following examples of the astronomical applications of photoelectric, thermoelectric and radioelectric phenomena will suffice to indicate that notable results have already been achieved by these new methods.

The photoelectric effect consists, as is well known, in the emission of electrons from a film of metal (calcium, for example) when it is illuminated by a beam of light: if the energy of the photons absorbed by the metal is sufficient (i.e. if the wavelength of the radiation is sufficiently short), they will dislodge the outermost electrons of the metallic atoms, and these will then be ejected to the exterior. We need not go into the details of this process (for which the reader is referred to any textbook on physics) but will merely recall that a photoelectric cell consists of an evacuated glass or quartz bulb on whose inner surface a film of some specially sensitised alkaline

Fig. 766 — REBER'S ORIGINAL RADIO-TELESCOPE

Fig. 767 — THE JODRELL BANK RADIO-TELESCOPE, THE LARGEST OF ITS TYPE IN THE WORLD. This picture gives an impression of the telescope's complex instrumentation. *(Keystone photo)*

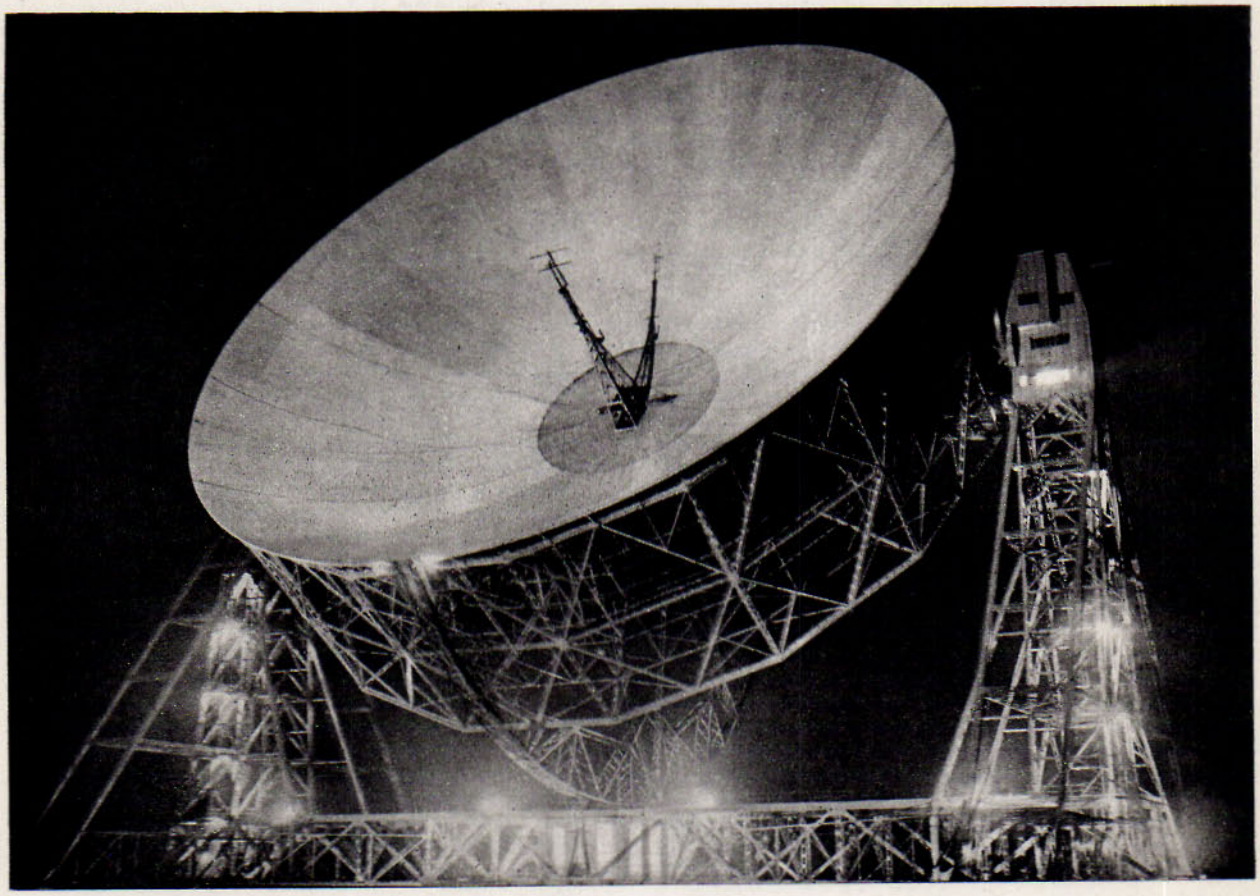


Fig. 768 — THE JODRELL BANK TELESCOPE IN OPERATION. The instrument is here shown tracking the signals from a Russian artificial satellite. *(Associated Press photo)*

metal has been deposited — usually sodium potassium or caesium; this constitutes the cathode (negative electrode). In front of it is mounted a wire or metal grid — the positive electrode or anode — whose function is to collect the electrons emitted by the cathode when light falls on it. A potential difference of about 100 volts applied to the two electrodes enables the electrons to cross the gap between them. The electrical charge collected by the anode was originally measured with an electrometer; this, however, is a very sensitive and delicate instrument and is therefore ill adapted to use in a portable installation mounted on an equatorial. Nowadays the current is more commonly measured by a sensitive galvanometer after passing through an amplifier of a type similar to those employed in a broadcasting station but employing special and very carefully constructed valves (Fig. 760). The electron-multiplier — a fairly recent device which consists of a cell with its own innate amplifying tendency — is in its turn rapidly replacing the external amplifiers previously used.

The photoelectric current is, within certain limits, directly proportional to the intensity of the light illuminating the cathode; this gives it an immediate advantage over the photographic plate, whose darkening is related to the intensity of the incident light in a very complex and variable manner. Moreover, the galvanometer readings give the required data without delay, so that in case of error or uncertainty the measurement can be repeated; photographic determinations, on the other hand involve a whole series of ancillary operations which generally involve several hours' work. However, the photographic plate possesses certain characteristic advantages, not shared by the photoelectric cell; these have already been referred to — permanent record of the observation, cumulative nature of the image, and the possibility of the simultaneous and wholesale recording of very large numbers of star images.

Up to the present time the photocell has therefore chiefly been used in astronomy to replace the eye in microphotometers (producing the modification of the latter known as the recording microphotometer — Fig. 761), and for the very precise determination of the magnitudes of bright stars and, in particular, of variables; but used in conjunction with the great Californian telescopes it has already produced valuable results in the investigation of spiral nebulae and the Milky Way (see Chapters 16 and 17).

The accuracy of photoelectric measurements is so great — often 10 times that of photographic — that they have been able to reveal such extremely delicate effects as stellar limb-darkening (detected and measured by Kron at Lick Observatory). That the photoelectric reaction is virtually instantaneous also makes it possible to analyse very rapid phenomena such as the change of brightness of a star at the moment of its occultation by the Moon; the comparison of such light curves (obtained, for example, by Whitford at Mount Wilson) with the curves calculated from diffraction theory has allowed the diameters of several stars to be estimated in this way.

Among the electronic devices already in use, notably in America, the following may be mentioned: an instrument which automatically counts the number of star images on a photographic plate, and a microphotometer which automatically traces the isophotal contours on a photographic image, simultaneously converting image-density to equivalent brightness. Many other developments of this kind are either planned or already in process of development. Among the latter, particular mention should be made of the 'electronic telescope', developed by A. Lallemand at the Paris Observatory; this is not, in fact, a telescope which uses electrons as a visual or photographic telescope uses photons, but an accessory instrument that converts the visible image formed by the optical telescope (refractor or reflector) into an electronic image whose available energy is greater than that of the visual image. Accordingly, astronomers now prefer to use the term 'image converter' for the device. This is achieved, in outline, as follows: the visual image falls on a thin photoelectric film in the focal plane of the instrument — this film being situated, as usual, on the inner surface of an evacuated glass or quartz tube; the electrons liberated from every point on the film are first greatly accelerated by the application of a difference of potential of some tens of thousands of volts and are then focused by means of external electric or magnetic fields (as in the electron microscope) on a photographic plate introduced at the other end of the tube (Fig. 762). A secondary electronic image, enlarged if necessary, is thus formed

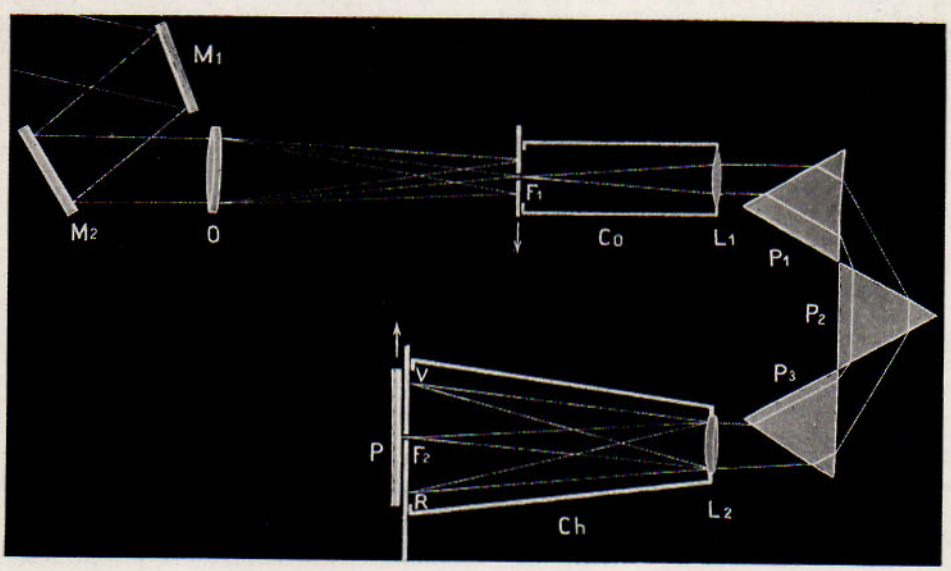


Fig. 769 — PRINCIPLE OF THE SPECTROHELIOGRAPH. The coelostat mirrors M_1 and M_2 feed light in to the objective O, which forms an image of the Sun at the entry slit F_1 of a monochromator; its exit slit F_2 isolates the chosen wavelength, which is recorded on the moving plate P.

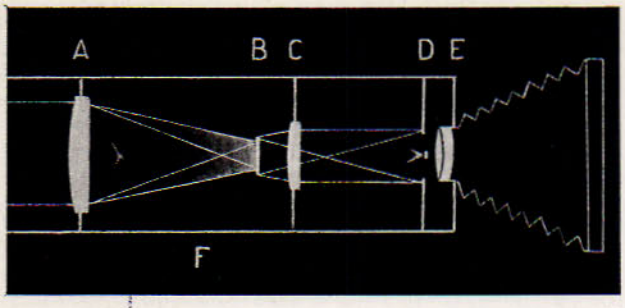


Fig. 770 — PRINCIPLE OF THE CORONOGRAPH. The simple lens A forms an image of the Sun on the disc B in its focal plane; an image of B is formed by lens C on the diaphragm D, whose function is to occlude diffracted light. The final image is formed by the photographic objective E.

on the photographic plate, and its point-to-point correspondence with the primary optical image reproduces the latter faithfully. The important feature of the apparatus, however, is that the energy available for the formation of its final photographic image has been greatly augmented (despite the weak output of the photocell) by the very considerable acceleration imparted to the electrons. In the final analysis it is thus a device for increasing light-grasp — or, more accurately, for increasing the 'speed' of photographic telescopes. Although the theory of this apparatus is simple enough, its practical realisation required much hard work and development before it could be effectively employed in astronomical research.

Nonetheless, as early as 1947 it was applied to the visual observation of variable stars by F. Lenouvel at the Haute-Provence station of the Paris Observatory.

Until recently, all these devices — the photoelectric cell, the electron multiplier tube and the image converter — were regarded primarily as photometric devices, since in the field opened up by Lenouvel's work they rendered particularly valuable — indeed, almost unmatched — service. In 1956, however, Lallemand succeeded in obtaining with his 'electronic telescope' some stellar spectra which encourage the hope that in the near future astronomy will have a very powerful new means of investigation at its disposal.

These spectra, which were obtained at the Haute-Provence national observatory with a 47-inch telescope, would have required an aperture of 275 inches had they been obtained with the orthodox photographic method. A transmission-grating spectrograph with 500 lines per mm. was used. Electronic photography produced images of an intensity that would have required exposures 20 to 25 times as long with the traditional photographic technique, using the most sensitive plates available. Furthermore, since the grains of the electronic plate are 100 times smaller than those in a photographic emulsion, very much finer detail can be recorded.

It can be seen from Figs. 763 and 764 that the electronic plate reveals a wealth of detail that is not recorded by the ordinary photographic plate; this is particularly obvious in the comparison spectrum (iron arc) which was exposed for 15 seconds in the latter case and only 1 second in the former.

Even better results are anticipated from a spectrograph now under construction, whose optical characteristics are adapted specially to this new observational technique. So long as ordinary photographic methods were employed, the need for such an instrument never made itself felt.

Already, stars of magnitude 14 can be reached with the 47-inch telescope, even used in conjunction with existing spectrographs.

The thermoelectric effect has also, during the last 40 years, made important contributions to the progress of astronomy. In essentials, the effect is this: if an electric circuit incorporates wires of different metals (e.g. iron and copper) welded end to end, and if one of these junctions is then made warmer or cooler than the other, a current will flow through the circuit; its strength (measured by a galvanometer inserted in the circuit) is proportional to the temperature difference of the two junctions. Thermocouples of this sort have many uses in industry, and they have been successfuly applied to astronomy by W. W. Coblentz at Lowell Observatory and by E. Pettit at Mount Wilson, using microscopic zinc-bismuth couples (whose elements measure only a few tenths of a millimetre) in a small evacuated tube mounted in the focal plane of a large telescope. If the image of a star is focused on one junction of such a circuit, the slight warming effect of the absorbed radiation is capable of producing measurable deviations of the galvanometer needle which are recorded photographically.

The essential superiority of the thermocouple over the photocell or the photographic plate lies in the fact that it is more or less equally sensitive to radiations of all wavelengths — in particular, it is sensitive to the long-wave infra-red radiations which produce no reaction in the photocell or the silver emulsion. And it is a consequence of the radiation laws (to be discussed in the next chapter) that the greater part of the energy emitted by a low-temperature body — below 1,500°, for instance — lies in this region of the spectrum.

It is primarily thanks to the thermocouple that the temperatures of the planets have been measured. The required region of the planet's image is focused on one of the junctions (Fig. 765) and the deviations of the galvanometer noted — first without any filter, then when various liquid and solid filters (water, rock salt, glass, fluorite, etc.) are inserted between the objective and the thermocouple, absorbing wider or narrower bands of the star's infra-red radiation. It is thus possible to differentiate, within the total radiation received from the planet, between the component consisting simply of reflected sunlight and that which is the real thermal radiation of the planet; almost all the former is confined within the visible and near infra-red regions of the spectrum (wavelengths less than 5μ), while the latter is situated almost entirely in the far infra-red — in the case of Mars, for example, between 8μ and 15μ, which is the limit of the transmission of the Earth's atmosphere. By comparing the galvanometer readings in these different cases, the quantities of energy within the various wavebands can be determined, and from this the temperature of the region in question can be calculated from the radiation laws (Chapter 20).

Thus the rôle of the thermocouple in astronomy is not the direct measurement of the temperature of a body (which would, indeed, be quite impossible) but the indication of the wavelength distribution of its radiation, from which the temperature can be derived theoretically.

Another thermoelectric instrument, the bolometer (developed by the American physicist, Langley), employs a different principle again, although it has points of resemblance with the thermocouple. In essentials it consists of a very fine platinum wire coated with lampblack, which is inserted in an electric circuit containing a battery and a galvanometer; in practice the platinum wire constitutes one of the arms of a Wheatstone bridge. If the wire is adjusted to lie parallel to the Fraunhofer lines and is moved mechanically along the length of the solar spectrum, the variations of its temperature resulting from changes in the quantity of the radiant energy that it

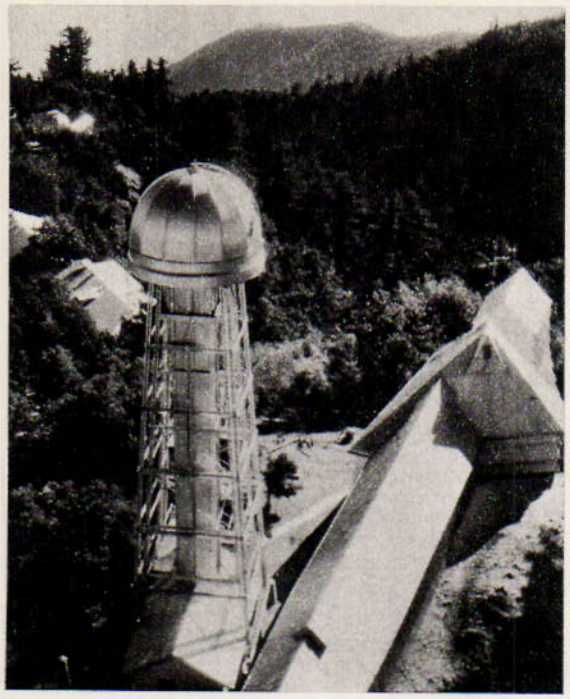


Fig. 771 — THE 65-FOOT TOWER TELESCOPE AT MOUNT WILSON OBSERVATORY, PHOTOGRAPHED FROM THE TOP OF THE 150-FOOT TOWER. In the lower righthand part of the photograph can be seen the horizontal Snow telescope of 24 inches aperture and 150 feet focal length.
(Photo, Wide World)

absorbs in different regions of the spectrum will produce corresponding changes in its electrical resistance and, therefore, in the current flowing through the circuit. It is only necessary to make a photographic record of the fluctuations of the galvanometer needle as the wire moves along the spectrum in order to have a faithful reproduction of the wavelength distribution of its energy.

Radio-astronomy and the artificial satellites. — Over the past 25 years radio waves have been added to the list of the radiations which we can employ in our study of the Universe. It is well known that the background noise of even the most perfectly constructed receiver can never be completely eliminated; a minimum noise level is imposed theoretically by the thermal agitation of the electrons in its valves, resistances and aerial, which produces a so-called thermal hiss. Now this theoretical minimum is itself never attainable in practice, even by the most highly perfected apparatus located far from every source of industrial or atmospheric interference. The origin of this inexplicable excess of background noise naturally became the subject of investigation.

K. G. Jansky, an American radio technician who applied himself to the problem of the irreducible background noise in short-wave receivers, was in 1933 led to the conclusion that part at least of the noise could be attributed to Hertzian waves of extraterrestrial origin. In about 1935 the employment of highly directional aerials (capable of picking up radiations from only a single direction) evinced the surprising fact that the intensity of this radiation increases as the aerial system is directed progressively nearer to the Milky Way, attaining its maximum value when the antenna is pointing towards Sagittarius — i.e. towards the galactic centre.

Jansky's experiments, which were carried out on the dekametre waveband, were repeated shortly before World War II on a shorter wavelength (1.87 metres) by another American, G. Reber of Wheaton, Illinois, whose apparatus was the first of the family of what are now referred to as 'radio-telescopes' (Fig. 766).

Since World War II, the development of this technique has gone ahead very rapidly indeed; the advances in radar which were necessitated by the war led to tremendous improvements in the study of short wavelengths, and these have been reflected in the progress of radio-astronomy. Here we shall only be able to review the outstanding aspects of this new branch of the science.

Certainly the most important contribution which it has made to astronomy is to render 'visible' the clouds of unexcited neutral hydrogen which make up such a great proportion of the matter of our Galaxy. It had long been regretted by astrophysicists that this extremely abundant element should not show up in its correct ratio. The characteristic spectrum of hydrogen does not appear in every spectrum observed throughout the Universe for the excellent reason that Lyman's series (No. 1) of hydrogen lines lies wholly in the unobservable ultra-violet, while Paschen's (No. 3) is situated in the far infra-red; only the Balmer series (No. 2) is visually observable, and this corresponds to a very restricted excitation range of from 10 to 12 electron-volts.

But in 1944 van de Hulst, of Leiden, predicted the existence of a particular neutral hydrogen emission that ought to be observable. Let us try to explain the principle of this emission as briefly as possible.

The proton and the orbital electron that make up a hydrogen atom behave like tiny magnets, and it is not a matter of indifference whether, within the magnetic field of the proton, the two parallel axes of the magnets lie in the same or in opposite directions. The switch of the electron's axis from one direction to the other may occur spontaneously, with the emission of a minimal quantum of energy corresponding to a wavelength of 21 cm. But this emission will, on the average, only be produced by a given atom once every 11 million years (the lifetime of the fundamental electronic state). In spectroscopic terminology, it is a 'forbidden' radiation. The probability of its occurrence is thus exceedingly small; if our chances of detecting it are not to be absolutely non-existent, the one wholly essential condition to be realised is that the hydrogen must be of extremely low density — for the frequency of collisions between its atoms decreases in proportion to the density of the medium.

An example of such a state of affairs is in fact provided by the interstellar material, which contains on the average only two atoms of hydrogen per c.c., corresponding to a density of the order of 10^{-22} gm./c.c. On the other hand, the interstellar void is so vast that the number of hydrogen atoms populating it is still very great. It therefore appeared possible that the radiation predicted by van de Hulst might be detected in the clouds of cosmic material situated in interstellar space.

And indeed, in 1951, evidence for the existence of this radiation of neutral hydrogen was discovered simultaneously by a Dutch radio installation under the direction of Professor Oort, an American installation at Harvard College Observatory, and an Australian installation at Sydney. During the few years which have since elapsed, immense progress has been made in the study of galactic structure by means of this radiation, and some of the most recent results have been referred to in Chapter 16.

A further consideration enormously augments the interest of these observations. The galactic clouds which give birth to this 21-cm. radiation are not stationary, but partake of the general galactic rotation. Now it is known

that this rotation is of such a type that rotational velocities are inversely proportional to distance from the centre. The rotational velocity of the Earth, lying some 8,200 parsecs from the centre of the Galaxy, is of the order of 135 m.p.s. The Doppler effect corresponding to this differential rotation — whose characteristics, thanks to the work of Oort and Lindblad, are well known — should therefore cause a detectable shift of the 21-cm. hydrogen line. This is in fact found to be the case, and it is by this means that we can deduce (from the velocities attributable to the various clouds emitting this radiation) their relative positions within the Galaxy.

The instruments used in this work are regularly being improved, both as regards 'light-grasp' and accuracy. In 1956 a giant radio-telescope went into operation at Dwingeloo in Holland; an interferometric antennal grid of large horizontal extent at the National Observatory in Haute-Provence, France, and another at Sydney in Australia. All previous radio-telescopes have since been overshadowed by the Jodrell Bank telescope in England.

Mention of this magnificent instrument should remind us that it is not only the receiving aspect of radio equipment which is employed in astronomy. Its normal transmitter-receiver function is also exploited. Three such uses should be referred to.

The first was the establishment of radar contact with the Moon by an American army station: the radio waves emitted by an aerial directed at the rising Moon were returned to the transmitter end of their journey, after reflection from the Moon's surface, 2½ seconds later — the time required for the distance at a velocity of 186,000 m.p.s. — and were there recorded on the screen of a cathode ray oscillograph. At present this achievement is little more than a curiosity, but it may have important technical and scientific applications in the future.

The second application which we must mention is the observation of the ionised trails of meteors, the principle of which (similar to that of ionospheric soundings) was explained in Chapter 9.

One of the most remarkable successes of this new technique is the observation of meteors in the daytime, for neither sunlight nor an overcast sky interfere in any way with the reception of the radio echo. It was in this way that the team of workers at Manchester under the direction of Professor (then Dr) A.C.B. Lovell were enabled in May-June 1947 to detect the extremely active radiant in the constellation Pisces. Because of its position close to the Sun its existence had until that time gone completely unsuspected.

The third application to be considered brings us again to the Jodrell Bank telescope, and to one of the newest of all aids to astronomy. Although the Jodrell Bank instrument was not designed as a 'super radar set', that in effect was what it became in the early days of artificial satellite observation, for it proved to be effectively the only instrument in the world capable of securing accurate data on the position of satellites whose radio transmitters had failed. This task it performed in the case of Sputnik II, supplying invaluable information.

Astronomy from Rockets and Satellites. — Although great advances in our knowledge of the Universe have been made in recent times by the application of modern techniques of optics and radio, two factors remain to hamper our progress. The first is the effect of the atmosphere in distorting and absorbing many of the waves reaching our instruments at the surface of the Earth, and the second is the vast distances separating us from the objects of our study.

The launching since October 1957 of a series of artificial satellites and lunar probes represents the first steps

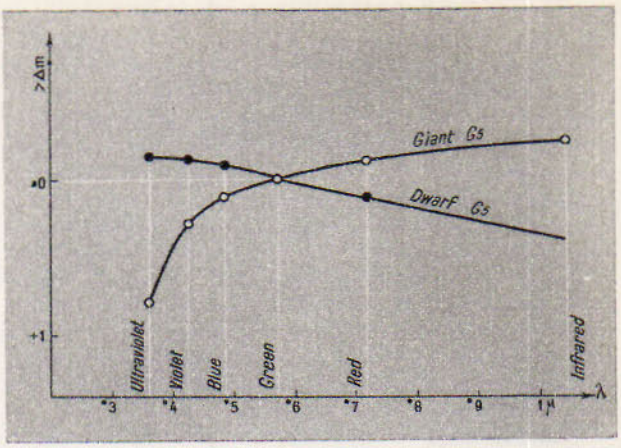


Fig. 772 — SIX-COLOUR PHOTOMETRY APPLIED TO A G5 DWARF AND TO A GIANT OF THE SAME SPECTRAL TYPE

in the attempt to overcome these limitations, to make observations without the distortions produced by refraction and turbulence in the atmosphere, and on those wavelengths which never reach the Earth's surface, and to carry out measurements actually in the regions studied.

The simplest of these experiments involves no more than placing in orbit a satellite moving at heights of a few hundred miles above the Earth. Observation of the motion of such an object yields data more accurate than can otherwise be determined of the shape of the Earth and the distribution of mass in the Earth, and also of the density of the Earth's upper atmosphere. Observations of the satellites already launched have shown that density at heights of 100—200 miles is several times greater than had previously been supposed, and also appears to fluctuate considerably, perhaps due to effects associated with solar flares and sunspot activity. It has been found that the flattening of the Earth is slightly less marked than was previously supposed, and that the northern hemisphere is flatter (by about 25 feet) than the southern.

But most of the knowledge we gain from satellites comes from the scientific experiments actually performed in the satellite during flight, the results of which are radioed back to Earth, either as they are produced, or first stored on a tape recorder or other instrument, and later relayed back to Earth when the satellite is over a suitable receiver.

These experiments fall into two classes, and each has produced results of great scientific value. The first class is designed to investigate the region of space in which the satellite or space probe is moving. The magnetic field may be measured, or the gas pressure, or the rate of impact with meteoric particles. So far the most important result has been the discovery of hitherto unknown belts of charged particles surrounding the Earth. These were found by Professor James Van Allen, of Iowa State University, with Geiger counters carried in the Explorer satellites launched in 1958, and the flight of two American space probes in 1959 added much new knowledge. From these it appears that the Earth is surrounded by two intense belts of radiation, one extending up to 3000 miles from the Earth, and the other three or four Earth's radii away. The question of whether these belts are permanent features, or vary with varying solar activity, or even the explosion of hydrogen bombs in the stratosphere, is of great interest, and is being studied by the 'paddle-wheel' satellite launched in August 1959. This satellite has an orbit which takes it between heights of 150 miles and 26,000 miles in a period of about 12 hours, and it passes through the Van Allen belts continuously.

All satellites launched so far have carried radio transmitters, both for tracking purposes, and to relay the information collected. But observation of the way the signal strength varies, quite apart from the information it carries, can give new knowledge of the ionosphere. This layer in the atmosphere which reflects radio waves of sufficiently low frequency is usually studied by means of an echo sounder, which transmits radio waves up from the ground and receives an echo back. As the frequency is increased from 1 megacycle upwards, the height from which the waves are reflected increases, greater concentrations of electrons being required to reflect the higher frequencies. At some critical frequency, which varies usually between 5 and 15 megacycles, the echoes cease, there being nowhere a sufficient electron density to reflect the waves. This gives a good picture of the ionosphere below the layer of maximum density, but can tell us nothing of what happens above this height. The study of the variation on signal strength received from a satellite as it moves through, or above, the ionosphere has yielded much new knowledge. It is now known that the number of electrons in the region above that of maximum electron density is several times the number below that height.

Thus the space immediately surrounding our planet is being explored, and in a similar way we shall soon know much more of our immediate neighbours in space, first the Moon, and then perhaps Venus and Mars. On the second day of 1959 a rocket launched from the U.S.S.R. passed 4,600 miles from the Moon and became the first artificial planet or asteroid in the solar system. This was followed by a similar lunar probe fired from the U.S.A. that also passed into orbit round the Sun, and in September 1959 by the first direct hit on the Moon's surface, scored by the Russian Lunik II. Since then both countries have pursued ambitious programmes.

But what of remote space? What can rockets and satellites tell us about the stars and distant galaxies? The future here seems bright, for it will surely not be long before telescopes can be mounted in satellites and take pictures of the universe using the great stretches of the spectrum that are absorbed in the atmosphere. For apart from the wavelengths of light to which our eyes are sensitive, from violet to red, and for radio waves with wavelengths between 1 cm. and 20 cm., the atmosphere is opaque. It is in effect covered with continuous and solid dark clouds. The shortest wavelengths, X-rays and ultra-violet light are absorbed high in the atmosphere, producing the ionosphere in the process. Then comes the narrow optical 'window', from violet to red, which penetrates the Earth's surface. Then the infra-red light, and the shortest radio waves penetrate the ionosphere, only to be absorbed rather lower in the atmosphere by various molecules, such as water vapour, oxygen or nitrogen. The long radio wavelengths cannot penetrate as they are reflected back into space by the upper side of the ionosphere.

Even in the optical range photographs taken in red and blue light give remarkably different pictures of parts of the universe, while on radio wavelengths the brightest objects are often very faint optically, and the brightest optical stars are not observable at all. Thus much may be confidently expected from pictures at other wavelengths obtained from satellites. Already Dr. Herbert Friedman, using rockets spending only a few minutes above the atmosphere has obtained maps of parts of the sky in ultra-violet light. These maps are comparable in resolution with those obtained by radio telescopes, and show some regions, the whole constellation of Orion for example, as great masses of brightly glowing material. This is a promising start to a new vital branch of astronomy.

But great as the possibilities for astronomical investigation opened up by the satellites may be, the fact remains that the enterprise of constructing and launching the rockets required to put even small instruments into orbit around the Earth is the most complex and expensive ever undertaken to make scientific observations. To illustrate this, some details follow of the facilities developed for the Vanguard satellites launched in 1958 and 1959 by the United States. A three-stage rocket was used for the launching of this and most other satellites so far put into orbit. The need for more than one stage lies in the fact that most of the initial weight must necessarily be fuel. When much of this fuel is burnt, a saving in weight can be achieved by separating the empty fuel tanks carried by the first stage, leaving a smaller second stage, and possibly a third stage to continue the flight.

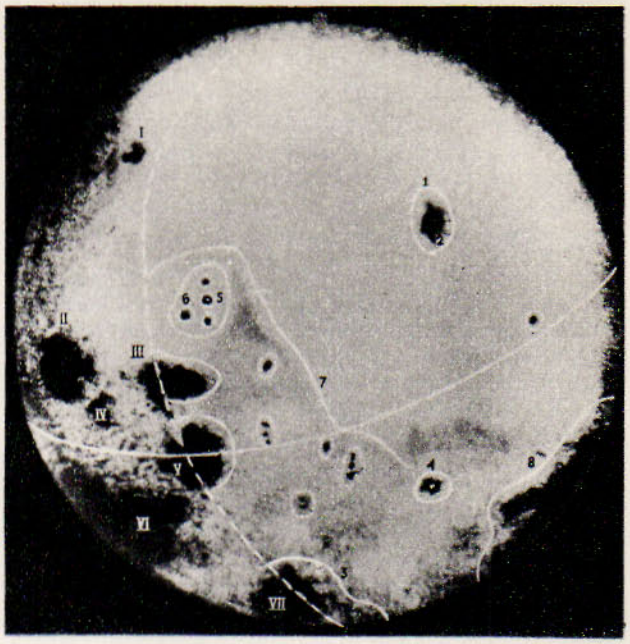


Fig. 773 — A PHOTOGRAPH OF THE OTHER SIDE OF THE MOON TRANSMITTED BY LUNIK III. The dotted line shows the edge of the visible part of the Moon. I: Humboldt Sea II: Sea of Crises III: Regional Sea IV: Sea of Waves V: Smith Sea VI: Sea of Fertility VII: South Sea. Newly revealed features: 1: Moscow Sea 2: Sea of Astronauts 3: Previously unknown section of the South Sea 4: Tsiolkovsky crater 5: Lomonosov crater 6: Joliot-Curie crater 7: Sovietsky mountain range 8: Sea of Dreams.

The Vanguard rocket was 72ft. long and 45 inches in diameter over the first stage, which constituted the lower 45ft. of its length. The total weight at take off was rather over one ton. The first stage when ignited, exerted a downwards thrust of about 1¼ tons, and this lifted the whole vehicle into the air. Two minutes and 22 seconds later the first stage burnt out at a height of 33 miles and with the rocket moving at a speed of 4,150 m.p.h. Immediately the second stage ignited and separated from the empty first stage, which fell into the sea 251 miles from the launching site at Cape Canaveral in Florida. Six minutes and eighteen seconds after launching the second stage was exhausted, the speed having reached 10,000 m. p. h. at a height of 122 miles. The rocket then coasted on under its own inertia until, when it was at a height of 261 miles and 762 miles out over the Atlantic the speed had dropped to 9,500 m.p.h. During this time a complex mechanism in the second stage ensured that the rocket was correctly aligned so that when the third stage was fired the satellite itself would go into orbit and not dive at great speed towards the Earth. Then the third stage

was given a spinning motion, to minimise any deflection produced by its burning, and was fired. Thirty seconds later the satellite was in orbit, 860 miles from the launching site, and moving with a speed of 17,000 m.p.h.

And all this was required in order to place a sphere 20 inches in diameter and weighing 21½ lb. into orbit. Not only this, but also a vast network of ground stations, equipped with radar, direction finding receivers, and receivers to collect the information transmitted from the satellite must be deployed and connected by telephone and radio with the control centre where electronic computers calculate the orbit into which the satellite is moving, and predict its future positions. More accurate positions of the Vanguards were obtained by a series of special Schmidt cameras, located round the world, and capable of recording the satellites' position to 2 seconds of an arc, and the time of observation to one thousandth of a second.

Launching a vehicle towards the Moon or into interplanetary space is similar, but instead of a velocity of 17,000 m.p.h. requited to put the satellite into a circular orbit round the Earth, at least 25,000 m.p.h. must be reached in order to give the vehicle sufficient energy to escape completely from the Earth's gravitational pull.

Because of the very great penalty in fuel and size of the launching rocket required for every additional pound of weight in the actual satellite, the power transmitted by the vehicle is limited, both because of the weight of the transmitter itself and, more important, because of the weight of the power supplies required. The use of solar cells, absorbing the radiation from the sun and converting this energy into electrical energy to drive the transmitter and other apparatus may give longer life to the instruments, but does little to reduce the total weight. Thus as the distance from the Earth increases the signals reaching the receivers become weaker and very great refinements in receiver design are required in order that the information transmitted from the satellite may not be lost. The American 'microlock' receiver does much towards solving this problem, increasing the effective sensitivity of the equipment a thousand times over a normal communications receiver, but even so, large aerial systems are needed to collect sufficient power to give clear signals. Although several such systems, or radio telescopes, of moderate size have been constructed specifically for this purpose, others designed for astronomical investigations have been called into service to follow the more distant probes that have so far been launched. The 250ft. radio telescope at Jodrell Bank has followed the three Pioneer rockets, two of which travelled about half way to the Moon and then fell back to Earth, and the third of which passed the Moon and went into planetary orbit round the Sun. The historic first flight actually to hit the Moon, by Lunik II launched from the U.S.S.R. in September 1959, was also tracked with this great instrument. As soon as it was known that the rocket was on its way, preparations were made to use the telescope on the appropriate frequencies, and a few hours later two receivers on frequencies of 20 Mc/s and 183 Mc/s were connected to the telescope. Details of the rocket's path were received from Moscow, and signals were received on both frequencies. On the second day of the rocket's flight the signals, although rather weaker, were again clearly received and followed until the moment of impact with the Moon. The frequency at 20 Mc/s was measured continuously with an accuracy of ½ cycle per second, and as the rocket approached the Moon this frequency fell, slowly at first, and then more rapidly as the Moon's gravitational pull accelerated the rocket away from the Earth. From the measured change in frequency the acceleration could be calculated, and at the end was exactly equal to that produced by gravity on the surface of the Moon, showing that a hit had been made, fair and square on the face of the Moon. Russian scientists soon announced that the signals received indicated no magnetic field or cloud of ionised gas similar to the Van Allen belts round the Moon.

Spectacular though the progress of the first years of space research may be, the results have posed at least as many new questions as they have answered old ones, and the satellites launched to date will be regarded in future as a rather crude start to a new mode of research, often hurried almost beyond endurance by the pressures of international competition and prestige. The near future will certainly bring new and better instruments to work on the present problems outlined above, and also new enterprises, some of which will make the achievements of the early satellites seem small by comparison. Already animals have been fired in rockets high above the atmosphere and retrieved safely, both in America and Russia. Already satellites have circled the Earth and been recovered. Already men have made space history with the first orbital flights round the Earth, the Russians Gagarin and Titov in 1961 and the Americans Glenn and Scott Carpenter in 1962. Soon rockets will probe further from the Earth, to Venus and Mars, and then beyond. Others will carry enough fuel to land instruments on the Moon without destroying them. One new experiment then possible would be the detection of tremors on the Moon, either from 'moonquakes' or from meteoric bombardment. And later man himself will move out into space, controlling his motion to the Moon and planets. But no date can be predicted for this with any certainty at present.

If this survey seems unfinished, that is as it should be. Space research is not yet six years old, and its techniques have not been fully established; not one of its possibilities has been fully explored and many remain untouched. Great advances will come with surprising rapidity.

The observation of radio sources. — The existence of radio sources in the Galaxy was established as long ago as 1931 by Jansky, in the United States, and it was soon discovered that the more intense of these are located in the galactic plane, and more especially in the direction of the constellation Sagittarius; the intensity of the latter source is comparable with that of the Sun. These results are more or less independent of the wavelength employed.

But very much more detailed investigations of the galactic emission have showed that there are a number of individual sources which produce a very sharp crochet in the intensity-trace when the Earth's rotation carries the source across the field of the (stationary) radio telescope.

Some of these deserve to be mentioned individually. One, discovered in Cygnus in 1948, can be identified with no visible star. On the other hand a source detected in Taurus coincides with the Crab nebula (pp. 333—334 and relevant figures), which is known to be the remnants of the supernova of 1054, mentioned in the Chinese records. A source in Cygnus has similarly been identified, quite recently, with a small nebula similar to the Crab; it lies very near the position in which Tycho Brahe in 1572 observed his supernova. Kepler's supernova of 1604 has likewise (1955) been identified as a radio source.

It appears probable that a particularly intense radiation emanating from a point in the constellation Virgo is to be attributed to an extragalactic source — a small, round extragalactic nebula in which an even smaller object is entangled. Modern opinion inclines towards the view that this is an instance of two extragalactic nebulae in collision; the same is true of a visually unidentified radio source in Cagnus, 200 million light years from our Galaxy.

Finally, the great extragalactic nebulae such as M 31 in Andromeda, and even certain galactic nebulae, have been identified as radio sources.

A list of 2,000 radio sources has recently been compiled at Cambridge by Ryle and his collaborators, the majority of which are situated in the northern hemisphere. All of these, with the exception of a hundred or so, are extra-galactic. The intensities of the weakest are only one-thousandth that of the Cygnus source; it is reasonable to suppose that the distances of these are some 30 times that of the latter, i.e. of the order of 6×10^9 light years.

The fact that the spatial density of these radio sources seems to increase with increasing distance is regarded as fitting much more naturally into the framework of curved and expanding space than into that of a static Universe.

Until a few years ago astronomers were afflicted by a partial but wholly aggravating blindness. Both the upper and lower ends of the range of electro-magnetic radiations which they could perceive, and were beginning to interpret, were cut off: the ultra-violet beyond 2900 A by the ozone in the Earth's atmosphere, and the infra-red beyond 20,000 A by the nature of their instruments.

With the aid of radio telescopes the astronomer can nowadays explore the decimetre-wave and metre-wave region of the electromagnetic gamut, while the use of rockets to carry instruments above the level of the atmosphere's ozone (15 miles approximately) has brought even more effective studies of the ultra-violet region.

This new harvest is proving to be exceedingly rich.

Special instruments for the observation of the Sun. — So far we have confined our survey of astronomical instruments to those whose usefulness is quite general. We must now describe some of the specialised instruments which have been designed for the study of the Sun, whose importance is so great — not just for astronomers, but for mankind at large — that whole observatories are dedicated solely to its investigation.

The most valuable, and also the oldest, of these instruments is the spectroheliograph, which was invented independently and almost simultaneously ba an American, Hale, and a Frenchman, Deslandres, in 1892. The purpose of this instrument is to provide monochromatic photographic images of the Sun — i.e. images formed with one only of the many radiations that comprise the solar spectrum; the most commonly used radiation is that of one of the hydrogen or ionised calcium lines (Chapter 10). The spectroheliograph consists in essentials (Fig. 769) of an objective which forms an image of the Sun on the entry slit of a monochromator — i.e. a spectrograph (see next chapter) which projects a spectrum on an opaque screen pierced by a second narrow slit, whose position can be adjusted to any required wavelength; a photographic plate placed behind this second slit will thus receive nothing but the radiation of the selected wavelength originating in the narrow strip of the Sun's disc covered by the first slit. By means of a mechanism for displacing the Sun's image uniformly across the first slit in accurate synchronisation with a movement of the photographic plate behind the second, a complete monochromatic image of the Sun is built up on the plate in a few minutes.

We have already considered some of the information regarding the structure of the chromosphere and prominences that has been provided by the spectroheliograph, and nothing more need be said regarding this. It may be mentioned, however, that a spectroheliograph is invariably a bulky piece of apparatus, several yards at least in length, and is installed either horizontally in a laboratory or in a vertical pit; in either case the Sun's light is fed into it by means of a heliostat — an arrangement of two plane mirrors mounted and driven in such a way that they always reflect the solar image towards the fixed objective. The horizontal layout, which is less costly than the vertical, has been adopted at Meudon Observatory; it suffers from the disadvantage that the prisms are exposed to troublesome variations of temperature which affect the steadiness of the spectrum relative to the second slit. The vertical layout, more complicated but more efficient, was adopted for the tower telescopes at the Mount Wilson and McMath-Hulbert Observatories in the United States; here the dispersive elements are maintained at a constant temperature at the bottom of a deep well, which the light traverses twice in passing from the heliostat to the plate. At the latter observatory (founded only about 20 years ago) the Sun's monochromatic image is recorded every day, continously and completely automatically, so as to provide a movie-film record of the Sun in monochromatic light. Such a luxury is not within the reach of most observatories, but a simpler, visual form of the instrument (the spectrohelioscope, evolved by Hale in 1924) permits the direct and continuous monochromatic surveillance of the Sun — an invaluable extension of the scope of the instrument when transitory phenomena such as flares are to be studied. Since Hale's invention of the spectrohelioscope the use of these instruments has become very widespread, and they are even to be found in the hands of amateur astronomers. The phenomenon of the persistence of vision is utilised to build up a steady image of part or the whole of the Sun's disc, which is scanned by the two slits sufficiently rapidly to eliminate any effect of flicker.

By means of the spectroheliograph or the spectrohelioscope the chromosphere and prominences can be observed at any time, but all attempts to observe the corona outside eclipse were unsuccessful until 1930, when the French astronomer B. Lyot, exploiting a totally different and very simple principle, was able to photograph the corona in full daylight. Lyot's coronagraph consists, in principle, of a simple refracting telescope, fitted with a circular screen in its focal plane; a system of lenses behind this screen forms an image of the Sun's vicinity and projects it either on to a photographic plate to give a direct photograph of the corona, or on to the slit of a spectrograph.

But the essential feature of the instrument — to which it owes its successful operation — consists in the elaborate and indispensable precautions that it embodies to eliminate all diffused light. Measurements of the brightness of the corona made during total eclipses have shown that it is about 10^6 times fainter than the photosphere; the brightness of the monochromatic radiations of its spectrum is perhaps 100 times less again. The least trace of haze in the Earth's atmosphere, a single speck of dust on the objective, or the smallest flaw in the latter (such as a bubble or scratch) is enough to scatter light whose brightness is equal to or greater than that of the corona itself.

The coronagraph must therefore be installed at a considerable height above sea level, and its object glass must be a *simple* lens, chosen for the perfect homogeneity of its glass, polished with the most scrupulous care, and efficiently protected against dust. Behind the blackened occulting disc, which covers the image of the Sun itself, is placed a lens which forms an image of the *objective* on a second diaphragm, whose function is to block all light diffused by the edge of the object glass; finally, a third lens projects the image of the occulting disc and its surroundings on to the photographic plate (Fig. 770).

Several coronographs are at present in operation: Lyot's original model at the Pic-du-Midi Observatory, two at the solar station of Harvard College Observatory at Climax (Colorado), at Sacramento Peak, New Mexico, at Arosa, in Switzerland, and others in Russia, Japan, etc. To give some idea of the coronograph's performance, it

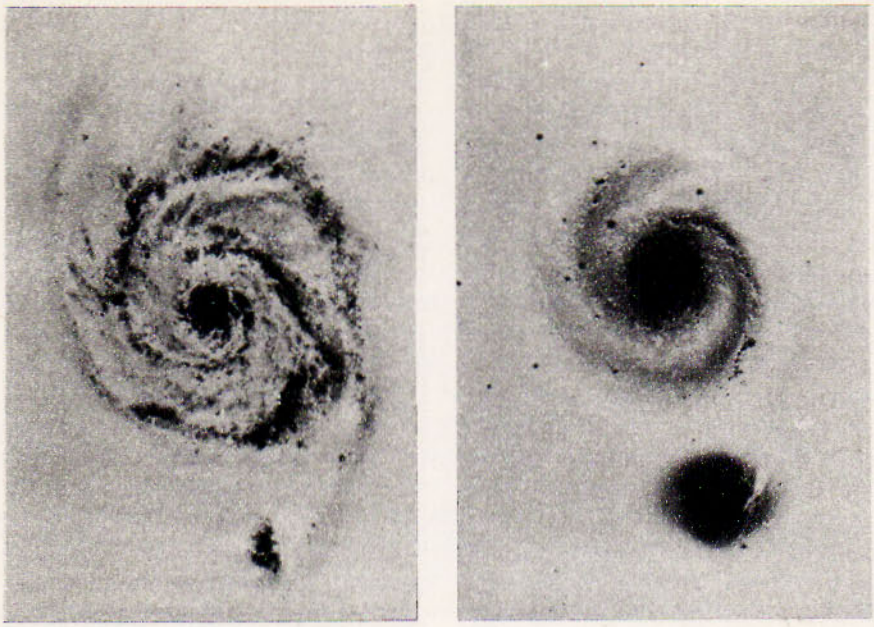


Fig. 774, 775 — ANALYTICAL PHOTOGRAPHS OF M 51, A SPIRAL NEBULA IN CANES VENATICI. Left: combination of yellow negative and blue positive; right: combination of yellow positive and blue negative. *(Zwicky)*

may be mentioned that with the Arosa instrument Waldemeier in 1941 succeeded in photographing the first magnitude star Regulus (α Leonis) at a distance of only 12′ from the Sun's limb.

The brightness of the blue sky, which still gives trouble even at high altitudes, is further reduced by the use of special filters which can isolate, more or less exclusively, one of the principal coronal wavelengths. It was in this connexion that Lyot and others developed and performed a new type of filter, some 20 years ago — the quartz monochromator, or monochromatic polarising filter. Although it is rather expensive, it is incomparably more effective than the usual coloured filters of glass or gelatin.

Lyot's earliest filter transmitted a spectral band which was 3Å wide. This band was centred on the most important green line of the coronal spectrum. A special device allowed it to be centred alternatively on the next most important line, in the red. The results obtained by Lyot, even with this initial filter — with it he was able to observe the corona in broad daylight — were sensational. Inevitably, however, they mainly concerned the inner corona, as distinguished from the fainter regions further from the Sun.

In 1948 Lyot perfected another filter, this time designed to permit the direct observation of the chromosphere in broad daylight. Its transmission band was only 0.75Å wide, and was centred on the Hα line of the hydrogen spectrum — this being of particular relevance in the observation of such chromospheric phenomena as prominences and flares.

It remained to develop a filter of this type that could cover a wider field, and this has been achieved by the two astronomers who took over the difficult responsibility of continuing Lyot's work: J. Rosch, director of the Pic-du-Midi Observatory, and A. Dollfus of Meudon Observatory.

An instrument capable of photographing the entire periphery of the Sun in monochromatic light was installed at the Pic-du-Midi in the spring of 1954. With its aid evidence was obtained for the first time of a connection between a centre of activity in the chromosphere and a directly observed coronal jet.

It has also been possible to double the selectivity of the filters by employing new combinations, thus greatly improving the conditions for coronal photography.

Similar advances have been made in the development of filters designed specially for the observation of the chromosphere (centred, as noted above, on the Hα line of hydrogen). A montage technique has been perfected which permits the continuous and automatic observation of the chromosphere. Such continuity of observation is vitally necessary because the flares — unlike spots and prominences (see Fig. 540) — are both sudden and short-lived, seldom lasting for more than an hour, and usually for very much less.

The general use of monochromatic filters in stellar observation. — We have seen the enormous service that has been rendered to the observation of the chromosphere and corona by Lyot's monochromators. Without going so far as to employ filters of such delicacy, perfection and narrow transmission bands, the modern astronomer nevertheless makes frequent use of coloured monochromatic filters in his studies of the stars. Such filters are valuable in a number of different fields, often being employed in conjunction with photographic emulsions of varying sensitivity ranges.

The reader will by now understand the distinction between the visual and the photographic magnitude of a star (pp. 304—305). The difference between these two quantities is known as the colour index of the star, and is characteristic of its spectroscopic type. Even if a star is too faint to yield a usable spectrogram its visual and photographic magnitudes can be determined, and its colour index will then be the basis on which it is allocated its position in the sequence of spectroscopic types.

It only became possible to exploit the concept of colour index fully when visual observations (which, necessarily, are always, to some extent subjective) were replaced by photovisual (p. 305). These are made with special photographic plates (isochromatic) which have been sensitised

to the red and yellow; they are used in conjunction with a yellow filter whose function is to reduce the intensity of the blue and violet. The spectral sensitivity curve of such emulsions is very similar to that of the human eye.

The emulsions used for photometric measurements — both isochromatic and orthochromatic (silver bromide/gelatine emulsions with a colour sensitiser) — have now been standardised by international convention. This makes it possible to compare the magnitude measurement of different observatories, obtained with varied equipment.

A preliminary generalisation of these techniques was undertaken with apparatus whose spectral sensitivity differs from those of the emulsions just mentioned. Other connotations of 'magnitude' have also been defined:

Photoelectric magnitude, as obtained with photocells of accurately known performance.

Radiometric magnitude, as obtained at ground level by means of instruments that react indiscriminately to all wavelengths, including the red and the infra-red (heat). Owing to the dependence of radiometric magnitudes on terrestrial atmospheric conditions and the reflectivity of the mirror, it has not yet been feasible to establish a precisely defined and internationally recognised standard.

Bolometric magnitude, which is a hypothetical measurement of the energy radiated by a star in its own vicinity; this likewise has not yet been given any precise definition which is accepted by international convention.

On the other hand some extremely interesting and satisfactory results have been obtained in recent years by Stebbins and Whitford, using coloured filters and photographic emulsion whose characteristics are defined by international convention. 'Two-colour photometry' is the determination of the magnitude difference when a star's brightness is measured, first, at a wavelength of about 4300 Å (isolated with a violet filter), and then at wavelengths around 5000 Å (with a yellow filter). The value so obtained is known as the 'international colour index'. All that this technique does is in fact to standardise and render more precise the values obtained by comparing (as has been done, very much more roughly, since the earliest days of astrophysics) the visual and photographic measurements of a star.

This technique has been extended to six-colour photometry, which allows the energy distribution curve of the star to be plotted as a function of wavelength. Such a curve is absolutely characteristic both of the spectral type and of the luminosity class (giant or dwarf) of the star in question. The giants of a given spectroscopic type are relatively very rich in infra-red, the dwarfs in ultra-violet. The data represented graphically in Fig. 772, refer to a giant and a dwarf of type G5 (similar, therefore, to the Sun). Their brightnesses have been made to coincide in the green region of their spectra.

Thus for each spectroscopic type and each luminosity class there is, in this internationally agreed system of calibration, a completely characteristic curve.

The same observational procedure has been applied to the galaxies. It has even, in the case of some of the nearest, been individually applied to the central and the peripheral regions (the former being much richer in infra-red). It is well known that the nuclei of the galaxies are in general rich in red giants and poor in blue. When, on the other hand, the curve is constructed for the galaxy as a whole, it reveals a superabundance of ultra-violet, owing to the very large numbers of blue giants in the spiral arms.

We can here do no more than remind the reader of the astonishing difference in the appearance of the night sky when it is photographed in blue light with an ordinary plate and in red light with a panchromatic plate and a red filter; the pair of photographs at Fig. 652 show that teeming myriads of stars escape our ordinary (in particular, visual) methods of observation.

These observational techniques have recently been greatly extended by the perfection of photographic emulsions sensitive to the infra-red as far as 8900 Å, which have been successfully applied to the investigation of interstellar matter, some of whose bright regions can only be observed in the near infra-red.

We must, however, deal at somewhat greater length with the sensational results obtained by F. Zwicky in regard to certain extragalactic nebulae which are sufficiently near for their detailed structure to be discerned. The technique employed was nothing more elaborate than black and white photography with monochromatic filters of differing spectral transmissions.

On a plate exposed through a blue filter, hot blue stars will record large images; cool, red stars, on the contrary, will have small images. (It must also be remembered that on a negative the images of stars are black spots, and that they appear as white spots only on the positive print.) Photographs taken through a red filter, conversely will record large images for the coolest stars (black on the negative, white on the positive), and small ones for the hottest.

It occurred to Zwicky to combine a positive of one kind with the negative of the other (C. Fehrenbach had also suggested a similar procedure, some years ago). If a blue negative and a red positive (transparency) are superimposed and printed together on the same plate, blue stars will appear as large black spots, and red stars as large white spots with small black centres. If, alternatively a blue positive is combined with a red negative, hot stars will be represented by white spots with small dark centres, while the images of cool stars will be dark spots. These effects are very obvious to the eye, and by suitably adjusting the colours of the filters, images can be obtained which allow stars of each family (cool or hot) to be readily identified even on a crowded plate.

This method of 'analytical photography' undoubtedly has a great future.

The first of the two analytical photographs of M 51 (Canes Venatici) reproduced here (Fig. 774) is a synthesis of a yellow negative and a blue positive. Hot stars register white, while low-temperature areas register dark — the large nucleus, for example, the satellite galaxy, and the spiral arms of cosmic gas, which here are very regular. On the outer edges of the spiral arms, on the other hand, can be seen indications of an abundance of hot supergiants, which have developed relatively recently from the diffuse material of these arms.

In the second photograph (Fig. 775) the combination is reversed: hot regions are dark, and relatively cool regions are light.* The nucleus and the satellite are less extensive than in the former photograph, showing that they are comparatively poor in very hot stars. The arms are bordered by a dark foam of very hot stars (blue supergiants), and the layout of the arms that they delineate is much more irregular than in the first photograph.

These two analytical photographs of M 51, obtained with the Mount Palomar telescope, should be compared with the very fine photograph (Fig. 677) which was taken with the Mount Wilson 100-inch in 1926 on an ordinary plate without a filter of any kind.

* The nucleus itself is an exception to this rule, appearing dark although it contains neither supergiants nor particularly high-temperature material of any sort. The reason is that it had to be over-exposed if the fainter arms were to be recorded.

CHAPTER 20

SPECTROSCOPIC ANALYSIS, AND ITS ASTRONOMICAL APPLICATIONS

Until the mid-nineteenth century it seemed that the fundamental aim of astronomy was the description and prediction of the motions of the heavenly bodies; this was, above all, the perfect field for the application of pure mechanics. The determination of the chemical and physical properties of the stars seemed so chimerical a project that the French philosopher Auguste Comte, in his *Philosophie positive* of 1825, had no hesitation in citing the chemical composition of the stars as an absolutely indisputable example of knowledge that mankind would never possess.

This rash pronouncement helps us to appreciate the enormous distance that astronomy has travelled during the past 100 years, thanks to the development of spectroscopic analysis and the discovery of the fundamental atomic and molecular radiation laws. In our survey of the general principles of these and of their more important applications, the exposition will be kept at the simplest possible level; nevertheless it may in places appear rather stiffer going than the rest of the book, but the reader will be more than recompensed for the small extra effort that this chapter may perhaps demand of him, by the increased understanding that it will give him of the mysterious but significant language of spectra, and of the amazing discoveries that modern astrophysics has extracted from their interpretation.

General principles of spectroscopic analysis, and astronomical spectroscopic instruments. — To start with, let us go back to the well-known and fundamental experiment performed by Newton in 1666: the decomposition of white light by a prism. If a ray of sunlight is allowed to enter a darkened room through a chink in the shutters and fall on a glass prism, it will be found that the ray is deviated on passing thourgh the prism and also (as can be seen if the emergent beam is intercepted by a screen — the wall opposite the window, for example) broken down to form a brightly coloured band; this consists of the familiar rainbow colours — violet, indigo, blue, green, yellow, orange, red. This band Newton termed the Sun's spectrum.

By passing single colours of the spectrum through a hole in the screen on to a second prism, Newton proved that these colours were themselves primary — i.e. were incapable of further decomposition. In 1800 William Herschel discovered, by moving a thermometer with a blackened bulb along the solar spectrum that beyond the red there are radiations (infra-red) which, like those of the visible spectrum, have a heating effect. And in the following year Ritter, a chemist, showed that invisible radiations likewise occur beyond the violet (ultra-violet), which resemble the blue and violet in that they are capable of exciting chemical effects; this he established by projecting the solar spectrum on to a sheet of paper impregnated with silver chloride, a chemical which is sensitive to light. From these experiments sprang the fallacious distinction between chemical, luminous and calorific radiations, which was current during the early nineteenth century. This distinction has been jettisoned by more recent physics, for the capacity of a radiation to affect the human retina or the photographic plate implies no special properties, and all these radiations, whether visible or invisible, are in fact of the same kind.

Before the opening of the nineteenth century the

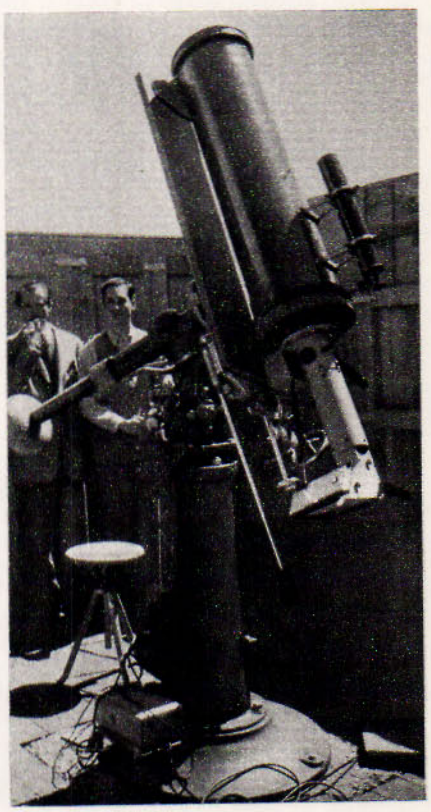


Fig. 776 — SMALL STELLAR SPECTROGRAPH mounted at the Cassegrain focus of a 10-inch telescope at the Jungfraujoch international scientific station, Switzerland. *(Photo, Chalonge)*

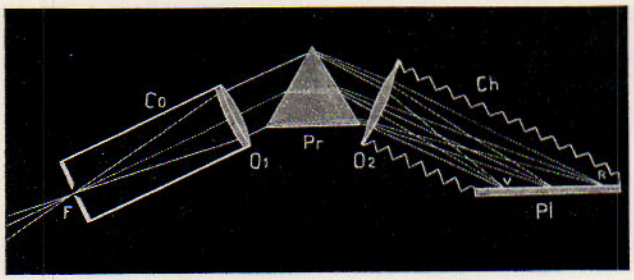


Fig. 777 — PRINCIPLE OF THE SLIT SPECTROGRAPH. Light from the telescope is focused on the slit F of the collimator Co, dispersed by the prism Pr, and collected in the photographic chamber Ch on the plate P1, where the wide-dispersion image of the spectrum is formed. The collimator objective O_1 which renders the rays parallel is achromatic; the objective O_2 may be a simple lens.

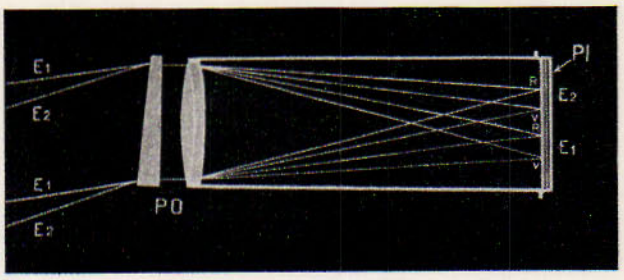


Fig. 778 — PRINCIPLE OF THE OBJECTIVE PRISM. The parallel beams from the stars E_2, E_1, etc., are deviated and dispersed by the small-angle prism P, mounted in front of the objective O. This produces in the focal plane (where the plate P1 is mounted) a low-dispersion spectrum of each star.

solar spectrum appeared to be undifferentiated — or, as it is called, continuous. But in 1802 the English physicist Wollaston found, on looking through a prism at a narrow slit illuminated by the Sun, that the spectrum's continuity was broken at seven points by transverse dark lines; the five most intense of these appeared to separate the principal colours quite distinctly from one another. In the same year the physicist Young laid the foundations of the wave theory of light, basing his conclusions on the phenomena of interference that he had discovered (see p. 460); he also accounted for the colours that are seen in thin films such as soap bubbles (which were a stumbling block to Newton's corpuscular theory of light), and finally provided a basis for the quantitative determination of the various colours of the spectrum.

According to Young's theory — whose essentials have survived the vicissitudes the theory of light has undergone since then — each colour, or more precisely, each primary monochromatic radiation, can be characterised by the period T of a vibratory disturbance, or by its frequency *vl*, which is the reciprocal of this period ($\nu' = \frac{1}{T}$). The period is not a directly measurable quantity, but in all types of propagation by vibration it is associated with a corresponding wavelength λ_o. If we put c for the velocity of light (299,776 k .p s., or 186, 300 m.p.s. *in vacuo*), then $\lambda_o = cT = \frac{c}{\nu'}$. In any medium other than air, whose refractive index is n, the velocity of light is reduced to $V = \frac{c}{n}$; accordingly the wavelength of the radiation becomes $\lambda = VT = \frac{V}{\nu'} = \frac{\lambda_o}{n}$.

The dispersive power of the prism is dependent on the fact that the refractive index of a medium is a function of the frequency of the radiation with which it is measured, and therefore of the radiation's wavelength; the diverse elements in a beam of mixed radiations will therefore be unequally deviated by the prism. This dependence of refractive index on wavelength is related in a complex manner to the structure of the medium, and does not permit absolute wavelengths to be derived from the observed deviations produced by the prism; some wavelengths must first be determined by an independent method.

This determination is made possible by the interference phenomena that occur when a ray of light is transmitted by a diffraction grating. Such gratings were in the early days constructed by winding a thin metal wire back and forth over the threads of parallel, very tight pitch screws; later, the technique of engraving extremely fine, equal and parallel lines on glass or metal with a diamond was developed. It is unnecessary to go into the theory of the diffraction grating, which is dealt with in any physics textbook; all that is necessary to know is that a ray of white light striking such a grating is split up into a series of diffraction spectra, which are said to be of the 1st, 2nd, 3rd . . . order, respectively; they are arranged symmetrically on either side of a central, undispersed and directly transmitted image (zero order). Also, if a is the separation of the lines of the grating, and y the wavelength of a given radiation, then Θ, the angles through which the images of the orders $p = 0, \pm 1, \pm 2, \ldots$, are deviated, is defined by the fundamental relation

$$a \sin \Theta = p\lambda$$

by means of which y can be accurately derived from direct measurement of p, Θ and a.

One advantage of the grating is that it provides a 'normal' spectrum, in which the deviation is very nearly proportional to wavelength, whereas in a prismatic spectrum the dispersion of the short wavelengths is much greater than that of the long; moreover, since the dispersion in diffraction spectra is a function of p, it is only necessary to use one of the higher order spectra (2 or 3) if greater dispersion is required. On the other hand, the prism has the advantage of concentrating all the available light into a single spectrum while the grating shares it out among the various images. Wood, an American physicist, has succeeded in producing gratings which concentrate 94% of the total incident energy into the 1st order spectra.

The wavelengths so determined are usually expressed either in microns (1μ = 0.001 mm) or angstroms (1Å = 0.0001 μ), the latter unit being named after the Swedish physicist who introduced its use. The values quoted in the tables are those measured in air (refractive index of about 1.0003), so that the wavelength *in vacuo* will be larger by about three ten-thousandths.

The earliest accurate determinations of wavelengths with a grating — and also the earliest detailed study of the solar spectrum — were made in 1815 by the Bavarian optician, Fraunhofer; using a small view telescope to observe the Sun's spectrum, which was formed by a prism mounted on the turntable of a goniometer, the light having first passed through a narrow slit located several yards away, he discovered several hundred dark lines crossing the continuous background of the spectrum (574 in all). The principal ones he designated by the letters A, B, C, . . . , H, K; a, b, . . . , h. This nomenclature has been retained, and the dark lines of solar spectrum are nowadays invariably known as the Fraunhofer lines.

The coloured plate facing p. 486 shows the principal

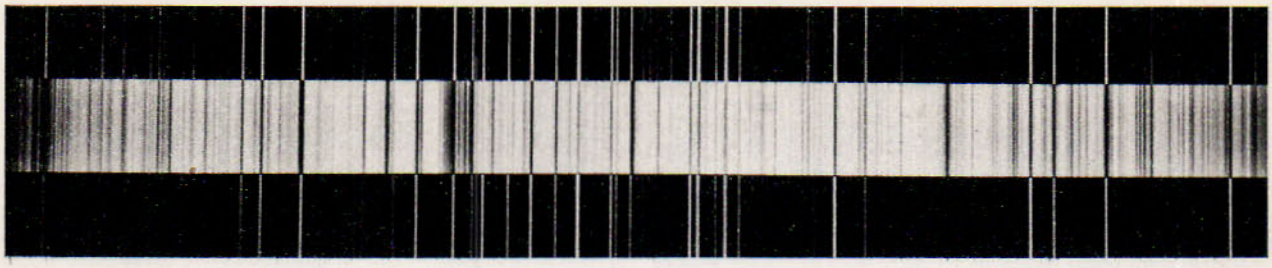


Fig. 779 — COMPARISON OF THE SUN'S SPECTRUM WITH THAT OF IRON. According to Kirchhoff's laws, the coincidence of the emission lines of the iron arc with many of the absorption lines of the solar spectrum as diffused by the Earth's atmosphere (the spectrum of the blue daylight sky) demonstrates the existence of iron vapour in the Sun's reversing layer. *(Photo, Lick Observatory)*

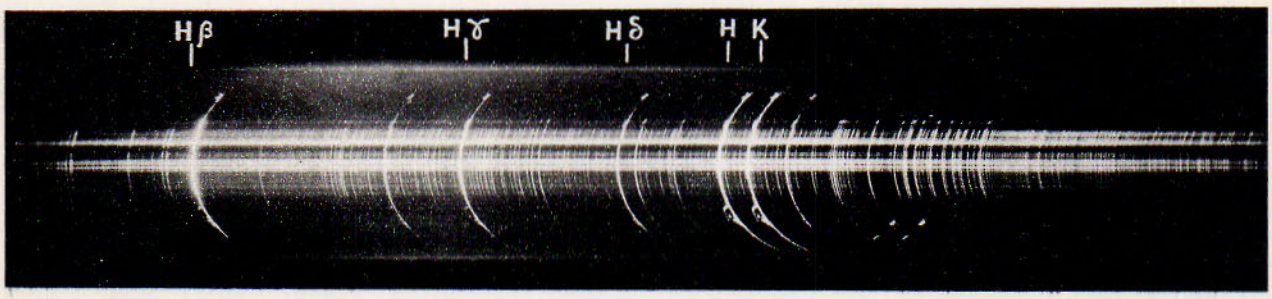


Fig. 780 — THE FLASH SPECTRUM PHOTOGRAPHED WITH AN OBJECTIVE PRISM. This spectrum, which appears for a few moment at the beginning and end of the total phase of a solar eclipse, is produced by incandescent vapours in the reversing layer, as explained by Kirchhoff's laws. *(Photo, Lick Observatory)*

Fraunhofer lines, and the following table lists their wavelengths (in air) and the elements producing them (the mechanism of this production will be explained later):

PRINCIPAL FRAUNHOFER LINES

LINE	WAVELENGTH (Å)	ELEMENT(S) RESPONSIBLE	ORIGIN
A	7584	Oxygen	Telluric band
a	7197	Water vapour	Telluric band
B	6867	Oxygen	Telluric band
C	6563	Hydrogen	H α solar line
D	5890, 5896	Sodium	Double solar line
E	5270	Iron, calcium	Solar lines
b	5167 – 5184	Magnesium, iron	Group of solar lines
F	4861	Hydrogen	H β solar line
g	4340	Hydrogen	H γ solar line
G	4308	Iron, calcium	Solar lines
h	4102	Hydrogen	H δ solar line
H	3968	Ionised calcium	Solar line
K	3934	Ionised calcium	Solar line

For the next 40 years or so after Fraunhofer's discovery, chemists and physicists occupied themselves with the study of a great number of spectra in the laboratory by means of the spectroscope, which assumed its final form in 1832 when it occurred to Babinet ro replace the distant slit by a collimator consisting of a very fine slit mounted in the focal plane of an objective; the latter therefore produced a virtual image of the slit at infinity. In 1843 the American physicist, W. Draper, constructed the first spectrograph, in which the solar spectrum was formed directly at the surface of a daguerreotype photographic plate.

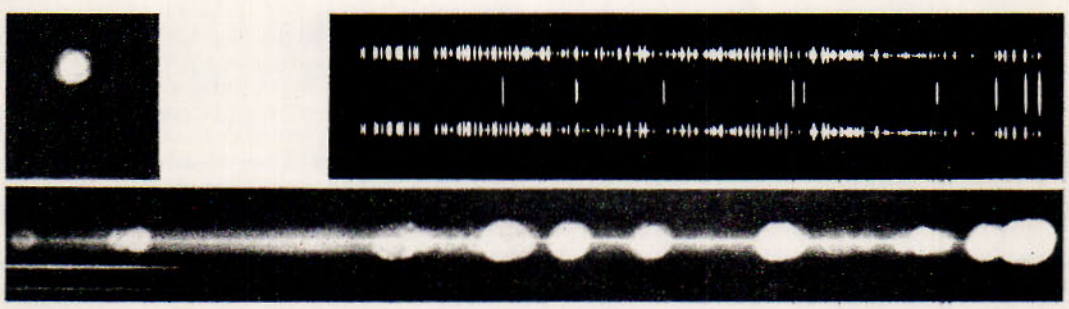


Fig. 781 — THE NEBULA NGC 7662 IN ANDROMEDA AND ITS SPECTRUM. Left, direct photograph with a 36-inch telescope (1914). Right, spectrum obtained with a slit spectrograph. Below, spectrum obtained with a slitless spectrograph. Note the different distributions of the various elements (hydrogen, helium, nitrogen, oxygen), or their different states of excitation, as shown by the appearance of the images in the lower spectrum and the unequal lengths of the lines in the upper. Extreme right, the doublet, N_1, N_2 of 'nebulium' (forbidden lines of O III). Compare with the spectrum reproduced in Fig. 638. *(Photos, W. H. Wright, Lick Observatory)*

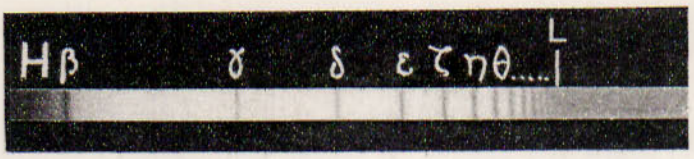


Fig. 782 — THE BALMER SERIES OF HYDROGEN LINES IN THE SPECTRUM OF VEGA, α LYRAE. *(Photo, Barbier and Chalonge, Jungfraujoch; 1935)*

These observations showed that a distinction had to be made between the continuous type of spectrum characteristic of incandescent solids and liquids*, and the spectra consisting of individual lines or bands that are produced by gases and vapours; the latter may in addition be either emission (bright lines on a dark background) or absorption (dark on a bright background). At that time, however, no general laws had been deduced regarding the production of spectra.

This somewhat confused state of affairs was clarified in 1862 by the investigations of the German physicist Kirchhoff, who in collaboration with the chemist Bunsen established the following facts: (1) every chemical substance has a specific and unique spectrum, (2) every substance is capable of absorbing the same radiations that it can emit, thus giving rise to the phenomenon known as the 'reversal' of spectral lines. In this way he accounted for the dark lines in the solar spectrum as being due to the absorption by the Sun's gaseous atmosphere of certain of the wavelengths in the continuous spectrum produced by the photosphere; the region of the atmosphere producing this absorption was accordingly termed the reversing layer.

Having established these two fundamental facts, Kirchhoff was in a position to go on and identify the lines of several of the chemical elements (hydrogen, iron, sodium, magnesium, nickel, calcium, etc.) represented in the solar spectrum: empirical spectroscopic analysis had been born. The new technique was next applied to other bodies by W. Huggins in England, Fr Secchi in Italy, and a host of others after them. By mounting a spectroscope at the focus of a telescope these pioneers created the science of astronomical spectroscopy and initiated the spectroscopic classification of the stars (see Chapter 14).

But the phase of really rapid development in astronomical spectroscopy had to await the application of the resources of photography; the first photograph of a stellar spectrum (that of Vega, showing four dark hydrogen lines) was obtained in 1872 by the American astronomer H. Draper, using a slitless, quartz-prism spectrograph in conjunction with a 28-inch telescope. But it was Huggins, in England, who made the earliest detailed photographic studies of the brighter stellar spectra, by fitting a spectroscope of normal type (i.e. incorporating a slit at the focus of an 18-inch telescope. The first spectrogram of Vega obtained with this equipment, in 1876, was already capable of revealing 7 absorption lines of hydrogen in the violet and near ultra-violet; the plate-holder being mounted in a frame that could be moved in a direction parallel to the spectral lines, Huggins was also able to record a comparison spectrum of the Sun (using sunlight reflected from the Moon or a planet) on the same plate. Wavelengths are nowadays calibrated by photographing, on either side of the astronomical spectrum, the spectrum of some well- known terrestrial source (iron arc, hydrogen discharge, etc.).

Finally, the American astronomer E. C. Pickering, remembering Secchi's idea of an objective prism, in 1882 set on foot the gigantic programme of stellar spectroscopic classification which in 1918 culminated in the publication of the *Henry Draper Catalogue*, which gives the spectral types of 225,000 stars. The operation of the objective prism relies on the fact that since the stars are effectively at infinity, their rays are parallel; a collimator is therefore unnecessary, and may be dispensed with, reducing the spectrograph to a prism and a photographic plate. This being so, it is only necessary to mount a large prism of small apex angle, in front of the objective of a camera in order to obtain low-dispersion spectrograms of all stars in the field that are bright enough to affect the emulsion. Since the monochromatic image of a star is a point, the resulting spectrum will not be a band but a fine line, the details of which will be very difficult to identify; the spectrum is therefore widened by using the slow motions of the equatorial to trail the image on the plate in a direction parallel to the edges of the prism; if the prism is oriented east-west, for example, the telescope must be shifted slightly in R. A. during the course of the exposure.

Objective prism, slitless spectrograph, normal spectrograph (with slit) — these are the basic instruments used in the spectroscopic analysis of the heavenly bodies. The first is adapted to the simultaneous study of large numbers of stellar spectra with weak dispersion, and therefore lends itself to survey and statistical work; the second is mainly used in conjunction with large telescopes for recording low-dispersion spectra of very faint objects which are beyond the reach of the objective prism; and the last, the classical instrument in spectroscopic analysis, is best suited for the detailed study of individual high-dispersion spectra of relatively bright objects.

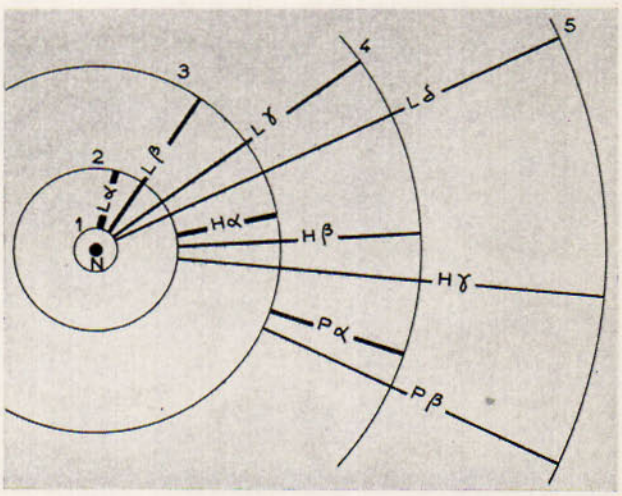


Fig. 783 — BOHR'S THEORETICAL EXPLANATION OF THE HYDROGEN SPECTRUM. Diagram showing the transitions between the various permissible orbits centred on the nucleus N, which give rise to the first emission or absorption lines of the Lyman series (Lα, Lβ, . . .), the Balmer series (Hα, Hβ, . . .), and the Paschen series (Pα, Pβ, . . .).

* It was later discovered that gases can also yield continuous spectra if they are subjected to sufficiently high pressures (as in the stars).

The atomic and molecular radiation laws. — While the earliest astronomical applications of empirical spectroscopic analysis were being developed, the physicists were striving to discover the laws regulating the emission and absorption of the monochromatic radiations which produce the lines of a spectrum. It was soon discovered that lines are produced by atoms, whilst bands (which, normally, are composed of numerous lines situated very close to one another) are characteristic of molecules.

It was also realised at an early date that a given element will produce different spectra according to its physical conditions, and in particular according to its temperature; thus were distinguished (in order of increasing temperature) flame spectra, arc spectra and spark spectra. But these differences long evaded physical interpretation.

The complexity of spectra likewise contributed to the difficulty of identifying the laws regulating the distribution of the lines, and the first simple empirical one was not in fact formulated till 1885, when the Swiss physicist Balmer found that the wavelengths of the then known lines of the hydrogen spectrum could be represented by a simple formula. If ν is the reciprocal of the wavelength, known as the wave number (i.e. the number of waves per cm.), R a constant whose value is 109,677 (Rydberg's constant), and n a whole number equal to 3, 4, 5, 6 for the successive hydrogen lines Hα (λ = 6563 Å), Hβ (4861 Å), Hγ (4340 Å), Hδ (4101 Å), then

$$\nu = \frac{1}{\lambda} = R\left(\frac{1}{2^2} - \frac{1}{n^2}\right)$$

As n is increased without limit, $1/n$ will tend towards 0, and the lines will therefore crowd closer and closer together as they converge on the limit of the series, which is situated at about 3645 Å in the ultra-violet. More than 30 lines of Balmer's series (emission and absorption) have been detected in spectrograms of the Sun and stars (Fig. 782). Later, Lyman discovered in the laboratory another series of hydrogen lines in the far ultra-violet (starting at 1216 Å), analogous to that of Balmer and expressible by a formula of the same type:

$$\nu = R\left(\frac{1}{1^2} - \frac{1}{n^2}\right)$$

where n = 2, 3 . . . Then Paschen discovered yet a third, in the infra-red, which can be represented by

$$\nu = R\left(\frac{1}{3^2} - \frac{1}{n^2}\right)$$

where n = 4, 5 . . . Futher series in the still more extreme infra-red have been detected by other physicists, all of which can be expressed by formulae of the same general type. Combining all these partial formulae, it appears that the complete hydrogen spectrum can be represented by a single simple expression:

$$\nu = \frac{1}{\lambda} = \frac{R}{n_1^2} - \frac{R}{n_2^2} \quad \ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where $n_1 = 1, 2, 3, \ldots$
$n_2 = n_1 + 1, n_1 + 2, n_1 + 3, \ldots$

The general form of the equation is thus $\nu = T_1 - T_2$, i.e. ν is equal to the difference of two terms of the same type, which are known as spectral terms. Later, the spectra of other elements were successfully represented by expressions which, though rather more complicated, were always of the same general type.

It was apparent that a general law had been uncovered and that this must reflect some fundamental property of atoms, but it was nevertheless a long time before a theoretical interpretation was found. It was not, indeed, until after the discovery of the electron structure of atoms* that the Danish physicist, Niels Bohr, in 1913 succeeded in evolving a satisfactory theoretical account of the monochromatic radiations and absorptions which produce the spectral lines.

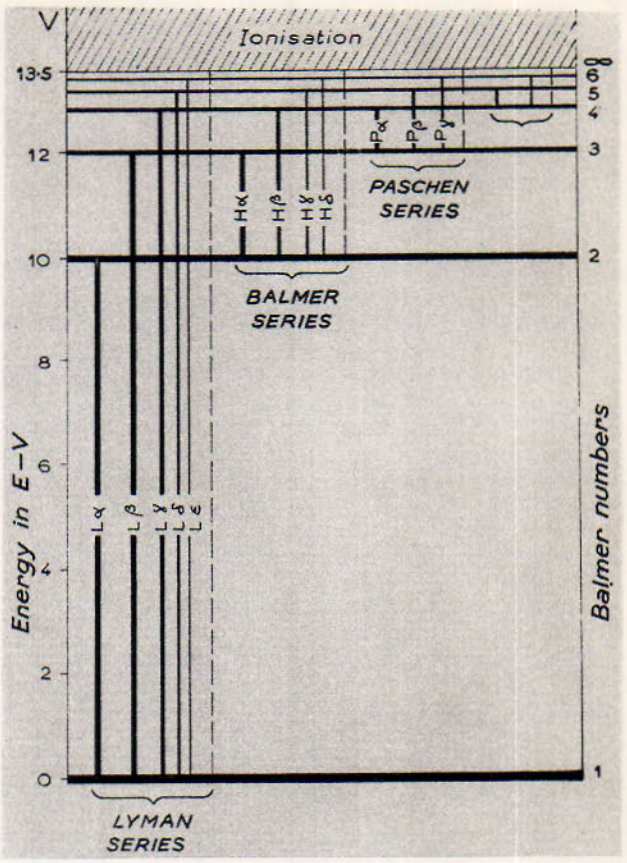


Fig. 784 — REPRESENTATION OF THE HYDROGEN SPECTRUM IN TERMS OF THE TRANSITIONS BETWEEN THE ATOM'S VARIOUS ENERGY STATES. Compare with the preceding figure. The thickness of the lines in the diagram is a rough indication of the relative importance of the transitions in each series and of the mean population of each level when the gas is raised to a high temperature. The energies are counted from the fundamental state of the unexcited atom. The Balmer numbers (which figure in his generalised equation) correspond to the different Bohr orbits.

The essential features of Bohr's theory are as follows. The electrons are not free to pursue any orbit whatever about the nucleus of an atom; only certain particular orbits (termed 'quantified') are stable, and the energies of electrons moving in these orbits under the influence of the electrical field of the nucleus have fixed and definite values, which are exact multiples of a fundamental constant; this constant had already made its appearance in the quantum theory of continuous spectra formulated by the Germany physicist Max Planck in 1900 (see below). Bohr then went on to assume that radiation is only emitted by an atom when an electron while circuiting the nucleus suddenly jumps from an outer to an inner one (Fig. 784), and that the frequency of the emitted radiation

* An atom may be likened, in some respects, to a miniature Solar System (with a diameter of the order of 1Å) consisting of a massive, positively charged nucleus, around which revolve a number of less massive negatively charged electrons.

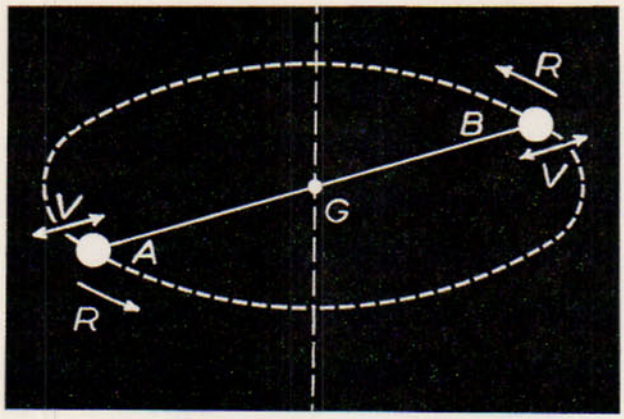


Fig. 785 — A MOLECULE CONSISTING OF TWO ATOMS A AND B. These may oscillate along the line AB (molecular vibration) whilst at the same time revolving about their centre of gravity G (molecular rotation). These motions are revealed by the structure of the bands in molecular spectra.

is proportional to the energy lost by the electron during this relapse towards the nucleus, the relation being expressed by the equation

$$h\nu' = E_2 - E_1 \quad \ldots\ldots\ldots\ldots \quad (2)$$

which gives the quantity of energy $h\nu$ (or quantum) which is emitted in the form of a photon (or 'atom' of light), as a function of the energies E_1 and E_2 of the electron before and after its change of orbit; h is a coefficient of proportionality known as Planck's constant, whose value is 6.6×10^{27} erg. secs.

Conversely, if an electron situated in its normal orbit nearest to the nucleus — which defines the fundamental state of the atom — receives a boost of energy from the exterior, either as the result of the atom's collision with one of its neighbours or of encountering a photon of appropriate energy, then it will jump to an orbit further from the nucleus, and in so doing will absorb a packet of energy in conformity with equation (2). Such is the fundamental mechanism of emission and absorption which underlines the formation of spectral lines; it provides a ready explanation of Kirchhoff's empirical laws.

However, an atom which has absorbed an excess of energy in the way just described does not as a rule remain for long in this abnormal, or excited, state: the electron almost invariably relapses after a very short interval (10^{-6} secs, or even less) into its original orbit. This it does either in a single jump thus liberating all its excess energy in one operation, or else in a series of stages via the intermediate orbits, when the energy is emitted in the form of several smaller quanta which therefore produce lines of longer wavelength, in obedience to equation (2).

Since the advent of Louis de Broglie's wave mechanics (1924), however — with its synthesis of the undulatory aspects (interference, diffraction) and the corpuscular aspects (quanta, photoelectric effect) of light — the outlines of Bohr's simple picture of electronic orbits have become somewhat blurred. At the same time Heisenberg's uncertainty principle (1927) has made the electron itself rather a vague entity, so that nowadays it is preferred to speak simply of transitions between different energy states of the atom, and to represent them diagrammatically as in Fig. 785. Often one is even content to speak simply of combinations of spectral terms, for if equation (2) is rewritten in the form

$$\nu = \frac{\nu'}{c} = \frac{E_1}{hc} - \frac{E_2}{hc} = T_1 - T_2^{*} \quad \ldots\ldots \quad (3)$$

the wave number ν of the radiation is freed from any pictorial association with electrons or orbits, and appears simply as the difference between the two terms T characterising different states of the atom.

But this new equation is highly reminiscent of equation (1) for the hydrogen spectrum: to complete the identity one has only to assume that $\frac{R}{n^2} = \frac{E}{hc}$, i.e. htat $E = \frac{Rhc}{n^2}$ represents the energy of the hydrogen atom when its single electron is revolving in the nth stable orbit. Now Bohr's theory allows the value of this energy to be calculated precisely, without any help from empirical data, and the outstanding success achieved by this theory was its derivation of a value for Rydberg's constant R (where $R = \frac{n^2 E}{hc}$) which agreed exactly with that obtained purely empirically by measuring the wavelengths of the hydrogen lines and applying equation (1). This agreement suggested most emphatically that Bohr's fundamental assumptions were correct; more recent developments have in fact confirmed the truth of his basic principles. These principles we must now outline, although we shall not be able to describe the later developments just referred to, since they lead at once to concepts of the utmost mathematical complexity when the theory is extended to atoms less simple than of hydrogen. Nevertheless, we shall have occasion to refer to some of them during the next few paragraphs. In any case, the important thing is to grasp the general principles.

The synthesis of the radiations corresponding to the various possible combinations of the spectral terms of an atom constitute spectrum I of the element in question; this is usually its flame spectrum. In fact, however, all the theoretically possible combinations of the terms do not, under ordinary laboratory conditions, result in the emission of radiation, since they are subject to certain theoretical restrictions characterising the relative probabilities of the different transitions that can occur within the atom. Thus an electron which undergoes a transition to a state of high excitation does not always relapse immediately to its original stable level of minimum energy: sometimes, when the transition probability of the lower level is very small, or if nothing occurs to dislodge the electron prematurely, it may remain for several seconds, minutes, or even longer, at certain 'metastable' levels, before relapsing finally to its original level and emitting the corresponding quantity of radiation. Under laboratory conditions the incessant collisions of the atoms with one another and with the walls of the containing vessel prohibit an atom from maintaining itself in a metastable state: it will give up its excess energy at the first collision, and will therefore relapse almost instantaneously to its original level. Thus a considerable number of radiations, corresponding to transitions from metastable states, have no chance of being emitted under laboratory conditions; because they are never encountered in laboratory spectra they were termed 'forbidden' lines.

In practice, also, the most usual spectrum of an element (its flame spectrum) consists of radiations which involve the outermost electrons almost exclusively, these being the ones most exposed to external influences. But under the stress of more vigorous thermal agitation (e.g. in the electric arc), one of these outermost electrons may escape

* Employing the usual conventions for expressing energies within the atom, one should rather write $\nu = T_1 - T_2$, with $T = \frac{-E}{hc}$ but this detail is ignored here in the interests of simplicity.

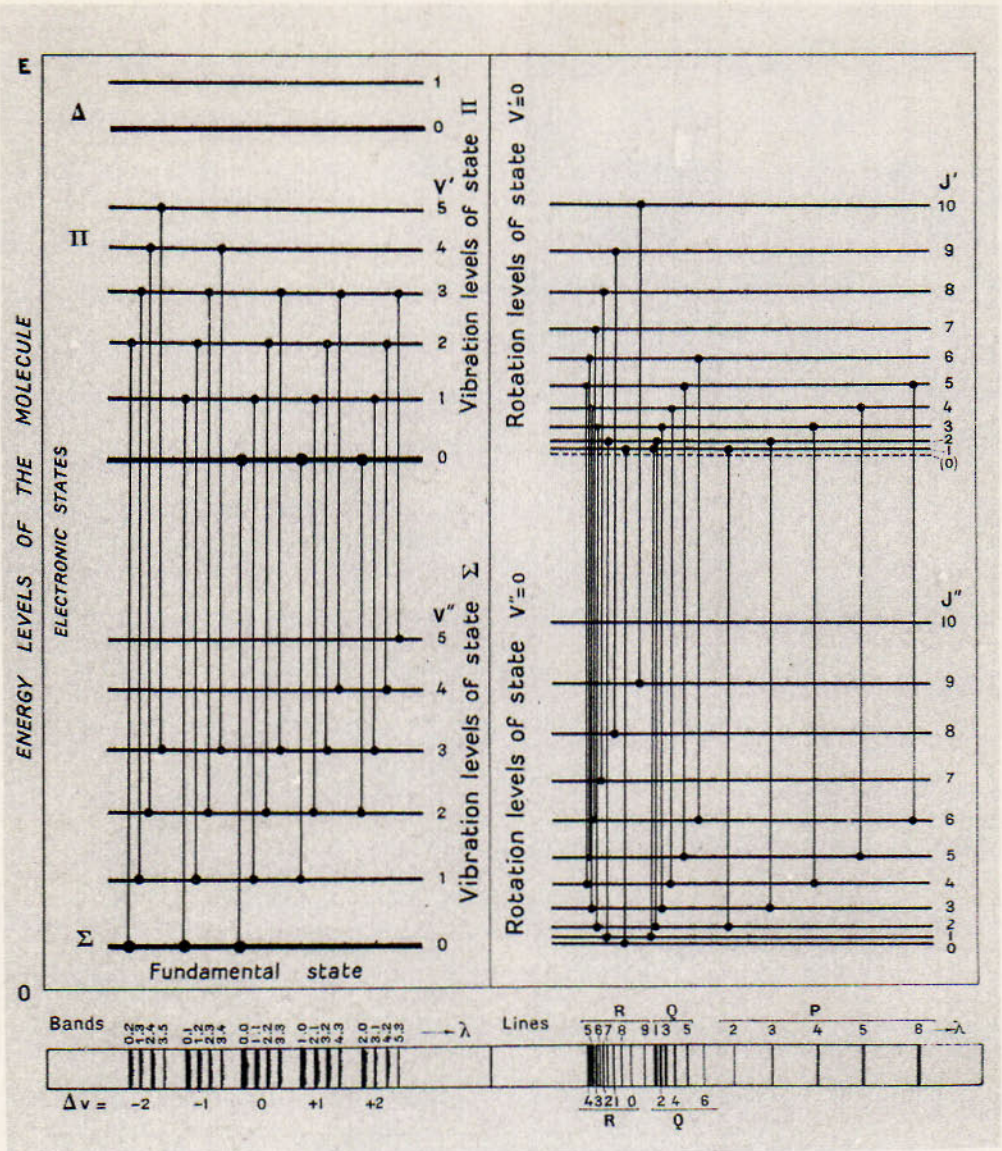


Fig. 786 — ENERGY LEVELS AND SPECTRUM OF A DIATOMIC MOLECULE. This diagram is the equivalent of Fig. 784 for atoms. Left, the first vibration levels, $v = 0, 1, 2 \ldots$, of the molecule's various electronic states Σ, Π, Δ, . . ., together with the transitions between the first two of these; these give rise to the 'sequences' of bands $\Delta v = 0, \pm 1, \pm 2, \ldots$ Right, the various rotation states for the first two vibration levels ($v = 0$) of the electronic states Σ and Π, together with the permissible transitions between them which give rise to the lines P ($\Delta J = -1$), Q ($\Delta J = 0$), R ($\Delta J = +1$), constituting the fine structure of the band (0, 0) which is generally the most intense. Each vibration level is subdivided again into a series of rotation levels responsible for the fine band structure observable in wide-dispersion spectra.

from the atom's sphere of influence altogether, being in fact ejected from the atom; this is the phenomenon known as ionisation. The arc spectrum (spectrum II) emitted by the remaining electrons of the ionised atom is very different from that of a neutral atom of the same element, since the loss of the outermost electron modifies its energy levels. Optically, an atom lacking one electron closely resembles a neutral atom of the element immediately preceding it in the periodic table of the elements, since the two have a similar electronic structure. For this reason, certain helium lines discovered by Pickering around 1900 in the spectra of very hot type stars were wrongly attributed to hydrogen; for helium, which in the neutral state has two electrons, after ionisation resembles neutral hydrogen in having only one. Should a hydrogen atom lose its single electron, nothing is left but the nucleus (proton), and it can therefore neither emit nor absorb radiation; it is thus completely invisible, and cannot be detected spectroscopically. The absence of its lines from the spectrum of a star does not by itself, therefore, amount to conclusive proof of the absence of the actual substance.

Subjected to still more violent disintegrating forces (e.g. the electric spark, collisions with greatly accelerated electrons, or, in the stars, very high temperatures which promote violent collisions with its neighbours or with already detached free electrons), the atom may lose a second electron; it then emits its number III (spark) spectrum. Continuing the same trend, a third electron may be lost, then a fourth, a fifth, and so on, until (if the

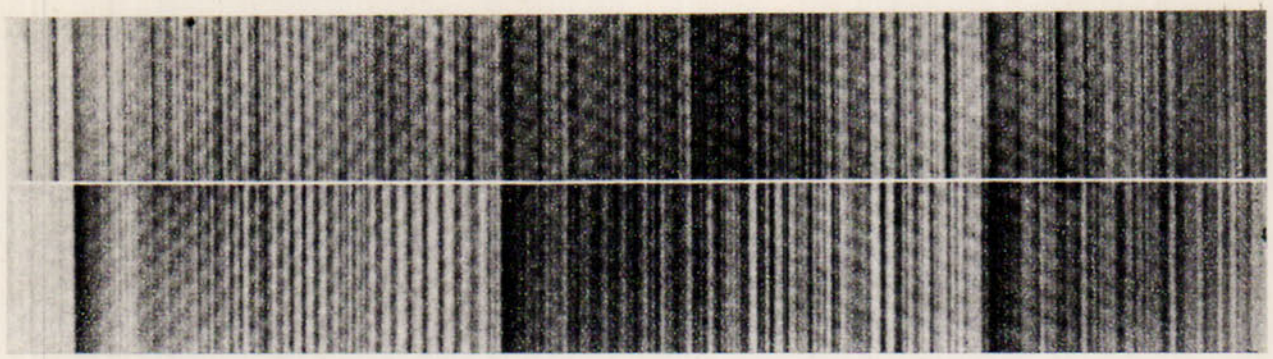


Fig. 787 — Above: titanium oxide bands between 7005 Å and 7150 Å in sunspot spectra; below: spectrum of the titanium arc in air. Except towards the head of each band, the individual rotational lines are resolved, thanks to the high dispersion employed. Compare with the righthand side of Fig. 787. *(Photo, G. E. Hale and W. S. Adams, Mount Wilson Observatory; 1907)*

physical conditions provide the requisite energy) it may have been bereft of a considerable proportion of its full complement of orbital electrons.*

Increasingly high energies are required to overcome the atom's resistance to this progressive stripping of its electronic cortege, for the very reason that the removal of negative electrons increases the relative attraction of the positive nucleus on those that remain**; for this reason the spectra of elements several times ionised become more and more difficult to obtain in the laboratory, whose resources (as regards the promotion of ionisation) are very soon exhausted. This difficulty is augmented by the fact that the positive ion which results from the ionisation of a netural atom strongly attracts any free electron in its vicinity; thus in a tube of gas that is being subjected to ionisation the abnormally high density of free electrons leads to their rapid recombination with the ions which therefore soon regain, more or less completely, their normal electronic structure.

This situation is completely changed if the atom is situated in a virtually boundless and extremely rarefied environment — as, for example, in interstellar space, a gaseous nebula, the solar corona, or even in the highest levels of the Sun's atmosphere. Here the singly ionised atom has a very small chance of encountering and capturing a free electron; a large number of atoms can therefore remain in even a highly ionised state for long periods, and are thus able to produce the spectra characteristic of this state with great intensity. For the same reason an atom will only very rarely come into collision with another, and will therefore have enough time to relapse spontaneously from a metastable state to its initial stable state, emitting a radiation which, in the laboratory, is 'forbidden'. It can thus be seen how it happens that familiar elements can sometimes emit, from astronomical environments, radiations whose spectra are wholly unknown in the laboratory: those of 'nebulium' (p. 369) and 'coronium' (p. 278) are examples which severely taxed the perspicacity of the physicists.

Such, in general outline, are the fundamental concepts of the theory of atomic spectra.

The theory of molecular spectra rests on similar foundations, but the complexity of these spectra is such that only molecules consisting of a small number of atoms have, up to the present time, been successfully investigated by the theorists. Here, therefore, we shall consider only molecules comprising two atoms — these are, in any case, much the most important astronomically.

The complex character of molecular spectra will at once be understood if it is reflected that in a diatomic molecule the motions of the electrons round their respective nuclei are superimposed on the motions of these nuclei relative to one another: the latter are, firstly, oscillations of distance on either side of their equilibrium position, constituting the 'vibration' of the molecule, and secondly, their rotation about a common centre of gravity (analogous to that of the two components of a binary star), constituting the 'rotation' of the molecule. Now these two motions are 'quantified', just as are those of the electrons: that is to say, only certain states are possible and these correspond to well-established frequencies and periods (of vibration and revolution respectively), i.e. to known energy levels, which are characteristic of the molecule in question. The emission or absorption associated with each spectral line invariably results from a transition between two energy levels in accordance with the fundamental equation (2).

For any given electronic state, however, the molecule may be in one of a whole series of different vibrational states, differing in amplitude, and for each of these it may be in one of another series of different rotational states, with different periods. Thus the molecule can produce

* The energy required to detach an electron from an atom or an ion is known as its 'ionisation potential', since it is commonly measured in terms of the potential difference that must be applied to a free electron to give it just sufficient kinetic energy to dislodge the orbital electron by collision. The atoms which are singly ionised most easily are those of the alkaline metals (sodium, potassium, etc.), which accounts for their photoelectric characteristics; a P. D. of a few volts is sufficient. The element most resistant to ionisation is helium, which requires a P. D. of 25 volts; it is for this reason that the lines of ionised helium (He II) appear only in the spectra of the hottest stars.

** For example, the successive ionisation potentials corresponding to the removal of 1, 2, 3, 4, ... electrons from the calcium atom are 6, 12, 51, 70, ... volts.

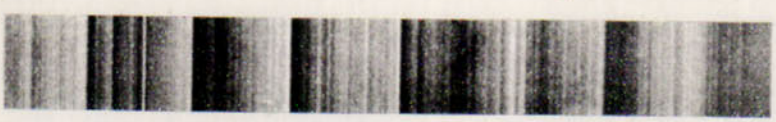


Fig. 788 — TITANIUM OXIDE BANDS IN THE SPECTRUM OF THE VARIABLE MIRA CETI (4600—7000 Å). The bands — unresolved, owing to the weak dispersion employed — produce the fluted effect characteristic of type M stars. Compare with the left-hand side of Fig. 786. *(Photo, V. M. Slipher, Lowell Observatory; 1907)*

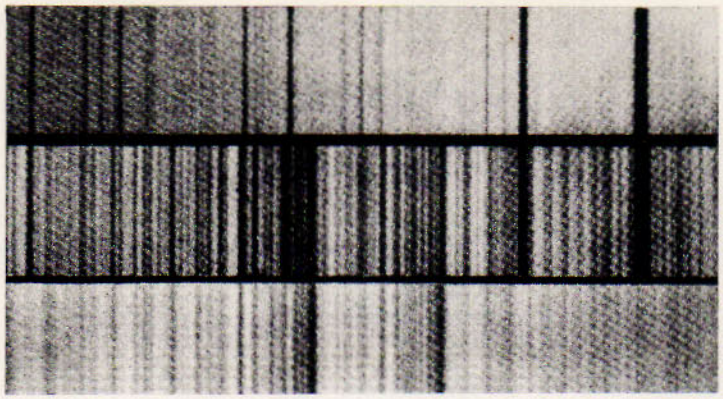


Fig. 789 — The spectrum of the Sun (top) compared with that of a sunspot (centre) and of a calcium hydride arc, 6365—6405 Å (bottom). The calcium hydride molecule, whose bands occur in the lower spectra, is dissociated at the higher temperatures of the photosphere. *(Photo, C. M. Olmsted, Mount Wilson Observatory; 1907)*

whole series of distinct spectra. If the transitions occur between two neighbouring rotational states (without affecting the molecule's electronic or vibrational states) the spectrum will be purely rotational, consisting of a series of sensibly equidistant lines lying, usually, in the far infra-red; these can be reproduced in the laboratory only with difficulty. If the transitions are between rotational levels associated with different vibrational states (the electronic condition of the molecule remaining constant), then a rotational-vibrational spectrum will be produced; this consists of a series of sensibly equidistant bands which may be resolved into individual lines and which normally occur in the near infra-red. It is to bands of this type (those of water vapour and carbon dioxide) that the strong absorption of this region of the solar and stellar spectra by the Earth's atmosphere is due.

Finally, if the rotational and vibrational levels between which transitions occur are associated with different electronic states, the molecule will produce its electronic spectrum, which lies in the visible region or the ultraviolet; it consists of a series of bands, just as an atomic spectrum consists of a series of lines. But whereas each electronic transition within an atom produced a single spectral line, the corresponding feature of a molecular spectrum is a series of bands resolvable into individual lines, each such band itself corresponding to the series of transitions between neighbouring rotational states (Fig. 786).

It is not difficult to conceive that such spectra will be of formidable complexity, and their analysis a matter of the utmost difficulty. Such analysis in fact only became possible after the discovery by the French astronomer Deslandres, in 1885, of the comparatively simple empirical formulae describing the distribution of lines within the bands, and of the bands within the whole spectrum. These formulae — which have since been justified theoretically — bear the same sort of relation to molecular spectra as those of Balmer bear to atomic spectra.

It frequently happens that the bands of a spectrum are arranged in a more or less regular sequence, each band presenting a characteristic appearance owing to the concentration of rotational lines in the region of a certain limit known as the head of the band, and their progressive spacing-out away from this limit. This produces, under low dispersions, an effect of the band being gradually shaded on one side (the 'wing') — whether on the short-wave or long-wave of the band depends on the nature of the molecule producing it. This 'fluted' appearance of molecular bands is an obvious feature of many astronomical spectra, particularly those of red, M type stars which contain strong absorption bands of titanium oxide, TiO.

The fine structure of these bands, produced by molecular rotation, can only be studied in the solar spectrum, since a degree of dispersion is demanded that the faint light of the stars will not support. It can be well seen, for example, in the A band of the solar spectrum (atmospheric oxygen), and also in the titanium oxide bands of spot spectra.

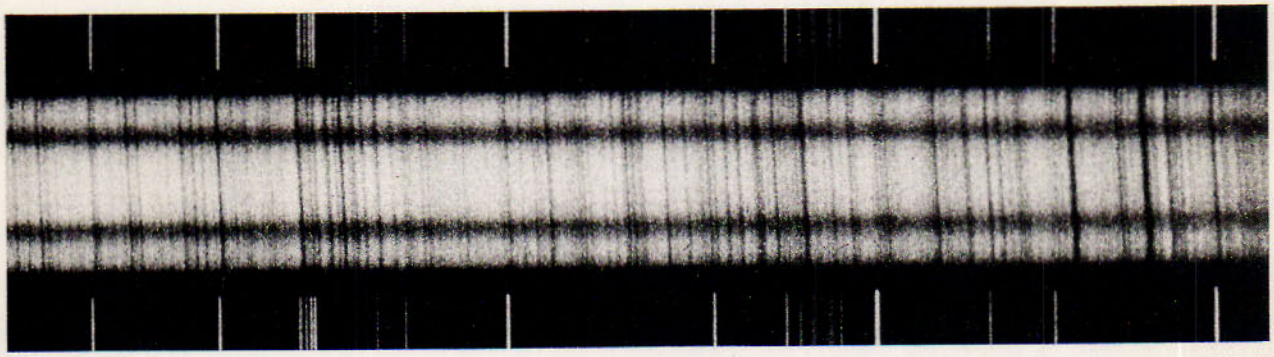


Fig. 790 — PART OF A HIGH-DISPERSION SPECTRUM OF SATURN AND ITS SYSTEM OF RINGS, obtained with the spectrographic slit oriented parallel to the major axis of the rings. From the inclination of the lines, the distribution of velocities across the globe and rings can be derived. Comparison spectrum: iron arc. This photograph also helps us to understand the mechanism of the widening of the lines in stellar spectra as a result of rotation; cf. Fig. 800. *(Lick Observatory)*

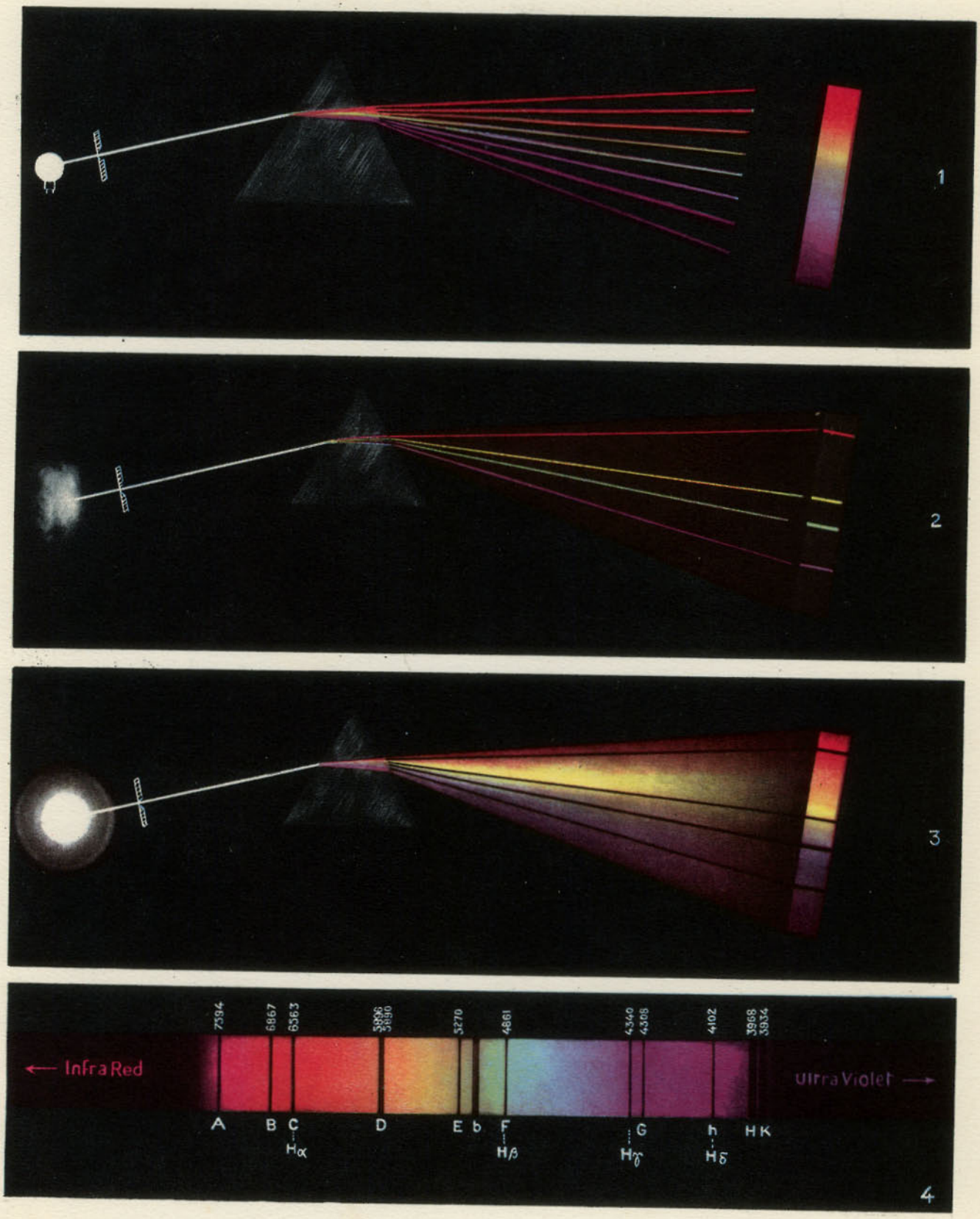


Plate XII. PRINCIPLES OF SPECTROSCOPIC ANALYSIS

1. Dispersion of white light by a prism. Continuous spectrum produced by an incandescent source, such as the filament of an electric light bulb or the Sun's photosphere.
2. Bright-line spectrum produced by an incandescent gas (e.g. mercury vapour lamp, Sun's chromosphere, etc.).
3. Continuous spectrum combined with a spectrum of absorption lines produced by an incandescent source enveloped by absorbing gas (e.g. solar photosphere and reversing layer).
4. Principal solar absorption lines (the Fraunhofer spectrum).

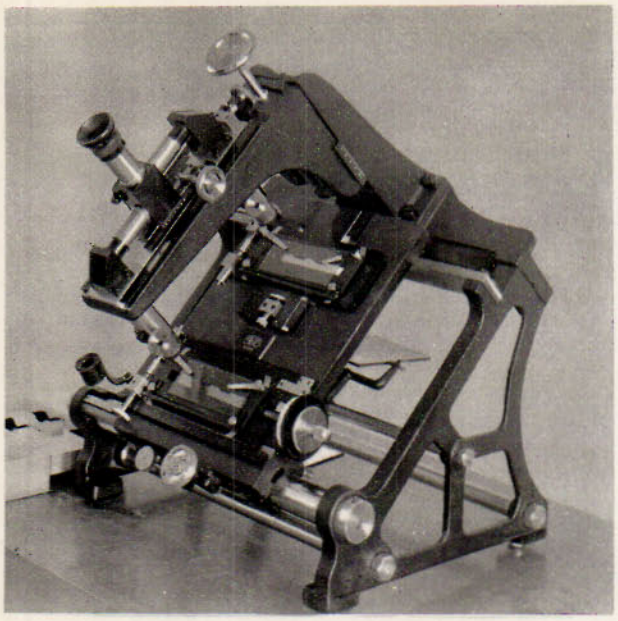


Fig. 791 – HARTMANN SPECTROCOMPARATOR, YERKES OBSERVATORY

Molecules are rather fragile combinations of atoms, and cannot withstand the violent thermal agitation produced by the high temperatures of some stars; only at relatively low temperatures can they escape dissociation into their constituent atoms. Bands of the most stable molecules – titanium oxide (TiO), carbon (C_2) and cyanogen (CN) – appear only weakly in the spectra of G type stars, whose surface temperature is about 5,000°. But they are abundant in spectra of K and M type stars (3,000° to 4,000°), and become so intense in those of R and N stars (2,000° to 3,000°) that they absorb almost the whole of the blue region, giving these stars their pronounced red colour.

To conclude, we have already seen that bands caused by various molecules have been identified in the spectra of low-temperature bodies such as the superior planets, comets, the atmospheres of Venus and the Earth, and even of interstellar space.

The Doppler effect, and its astronomical applications. – The reader has probably noticed that when he is driving fast in a car, and another car approaches and passes him with its horn sounding, the pitch of the horn suddenly falls as the car passes him. The same phenomenon is to be observed from the platform of a station through which a non-stop express is passing with its whistle going; and, in general, whenever the velocity of the source of a steady-pitch sound changes rapidly with relation to the observer. This effect – to which the Austrian physicist, Doppler, drew attention in 1842 – is simply explained by the fact that the sound waves reaching the ear are obviously more widely spaced from one another when the source of the sound is receding from the observer than when it is approaching him, owing to the displacement of the source between its emission of two successive vibrations. Thus the apparent wavelength of the vibration is shorter in the case of an approaching source than in that of one which is receding; this variation of wavelength manifests itself subjectively as a change in pitch – rising in the former case, falling in the latter.

It occurred to Doppler that an analogous effect ought to occur with light waves, and that therefore the observed colour of a star would depend on its motion towards or away from the observer. This idea is perfectly sound in principle but misleading when applied, since the velocities of the stars relative to the Earth are too small (a few hundreds of m.p.s. at most) compared with that of light (186,000 m.p.s.) to produce any perceptible modification of colour. Some years later the French physicist Fizeau showed that the effect might nevertheless be detected by

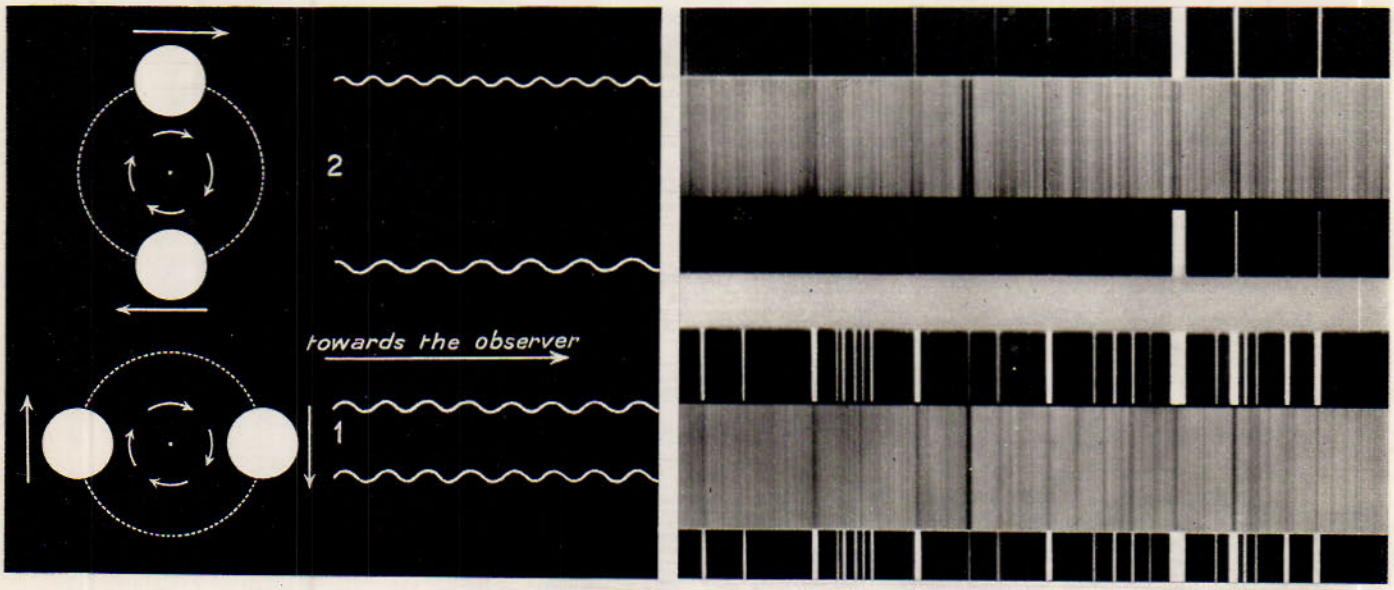


Fig. 792 – PRINCIPLE OF THE DOUBLING OF THE SPECTRAL LINES OF A SPECTROSCOPIC BINARY WITH SIMILAR COMPONENTS (DOPPLER EFFECT): Mizar, ζ Ursae Majoris. In *1*, the orbital velocities of the two components are perpendicular to the line of sight; their radial velocities are therefore equal, their radiations are of the same wavelength, and the lines of their spectra are superimposed. In *2*, a quarter of a revolution later, on the other hand, the velocity of one component is directed towards the observer along the line of sight and that of the other away from him; their radial velocities are therefore different, the wavelengths of their two sets of radiations will be slightly different, and the spectral lines of the two components will appear side by side in pairs, symmetrically disposed with respect to their earlier common position.

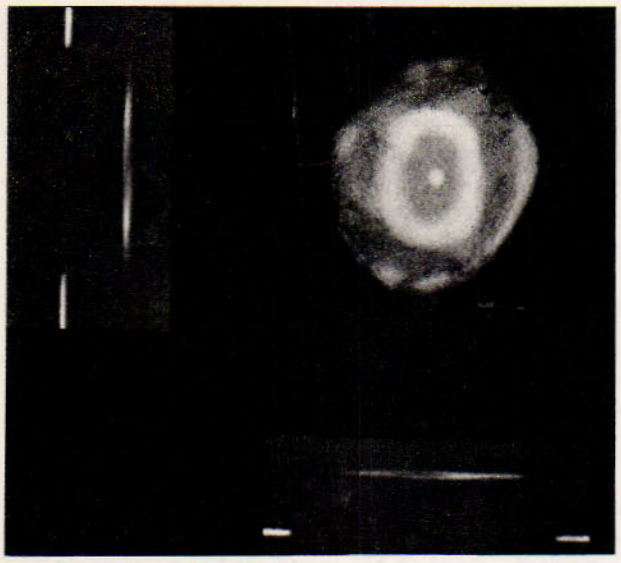


Fig. 793 — DOUBLING OF THE LINES OF THE PLANETARY NEBULA NGC 7662 IN ANDROMEDA, revealing the expansion of its gaseous envelope. The N_1 line of ionised oxygen ('nebulium') as observed when the slit is placed along the major axis (left) and the minor (below). Compare with the spectrum reproduced in Fig. 781. *(Lick Observatory)*

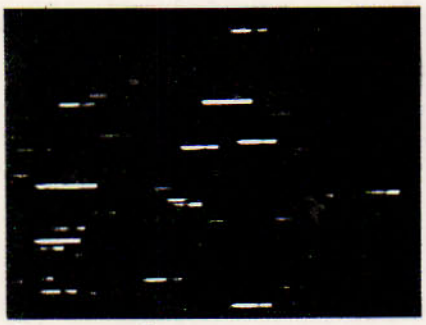


Fig. 794 — STELLAR SPECTRA OBSERVED WITH AN ORDINARY OBJECTIVE PRISM.

measuring the positions of the spectral lines of a star: these provide a very much more precise criterion than the tint of the star, and are therefore capable of betraying very much smaller relative velocities.

The expression for such velocity-shifts is

$$\lambda' = \lambda \left(1 + \frac{v}{c}\right) \ldots\ldots\ldots (4)$$

where λ is the true wavelength, as measured when the source and the observer are relatively at rest (e.g. in the laboratory); v is the apparent wavelength of the same-radiation when its source is in motion relative to the observer (i.e. when its radial velocity is not zero — see p. 307); and c is the velocity of light. This expression has been verified to a high degree of precision, in the laboratory.

Radial velocities are counted as positive if the source is receding from the observer, and negative if it is approaching. It is thus clear from equation (4) that a velocity of recession will produce an increase of wavelength, involving a displacement of the spectral lines towards the red end of the spectrum; a velocity of approach, on the contrary, will displace the lines towards the violet.

It should be noted that a motion of the observer towards the source will produce the same effect as that of the source's motion towards the observer: the operative factor is change of separation, and it makes no difference from which end the change, if any, is made.

Consider, for example, the displacement of the lines in the spectra of the stars produced by the Earth's annual revolution about the Sun; this displacement will be maximal in the case of stars situated on or near the ecliptic, such as Antares or Procyon. When the Earth is moving directly towards a star, at a velocity of 18.5 m.p.s., (i.e. when the star crosses the meridian towards the end of the night), the relative change in the wavelengths of its spectral lines will be, from equation (4),

$$\frac{\lambda' - \lambda}{\lambda} = \frac{v}{c} = \frac{-18.5}{186{,}000} = -\frac{1}{10{,}000}$$

In other words, the lines will be displaced towards the violet by an amount equal to one ten-thousandth of their normal wavelengths: thus the D lines of sodium (5890 Å and 5896 Å) will be displaced by about 0.6 Å, or one tenth of their separation. The accurate measurement of so small a shift demands the employment of a very high-dispersion spectrograph.

Six months later the Earth will be moving directly away from the star, which will now be on the meridian during the evening. The relative velocity will have

Fig. 795 — SPECTRA AND INVERTED SPECTRA OF STARS OBSERVED WITH AN OBJECTIVE PRISM BY FEHRENBACH'S METHOD. The displacement of equivalent lines permits a star's relative radial velocity to be calculated.

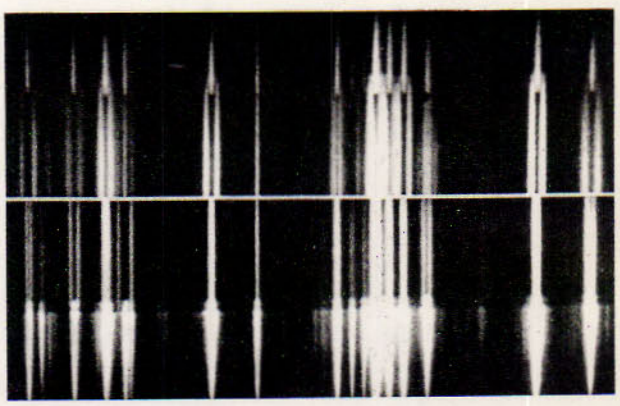


Fig. 796 — THE ZEEMAN EFFECT: the splitting of the lines of a spectrum under the influence of a magnetic field in the source. Doubling of the lines of titanium (4285 – 4310 Å) observed parallel to (above) and perpendicular to (below) the lines of force of an electromagnet. The outermost spectra (above and below) are comparison spectra of titanium in the absence of a magnetic field. *(Photo, A. S. King, Mount Wilson Observatory; 1909)*

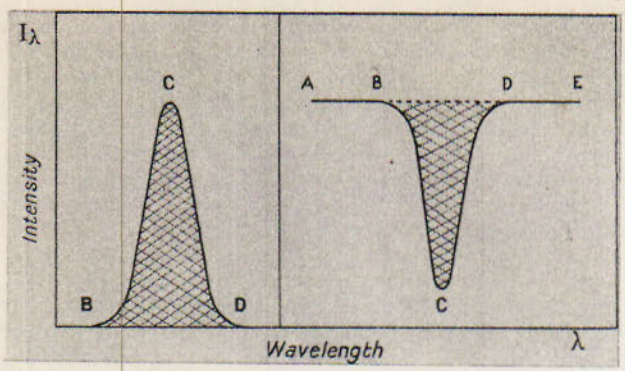


Fig. 797 — WIDTHS AND PROFILES OF SPECTRAL LINES. The curve representing the variation of the energy of the radiation as a function of wavelength determines the profile of the line. Left, an emission line; right, an absorption line. The shaded area delimited by these curves is a measure of the total energy emitted or absorbed by the line.

changed its sign, and the displacement of the spectral lines will be equal to that encountered in the previous case but in the opposite direction — i.e. towards the red. It will readily be seen that the mean of the two measurements will give the star's radial velocity relative to the Sun, and measures of stellar radial velocities are always corrected in this way in order to free them from the effects of the Earth's motion.

This is the principle of a recent and extremely ingenious technique for calculating the Sun's distance. First suggested by the English astronomer, Gill, towards the end of the nineteenth century, it has been successfully applied by various astronomers since 1905, and has produced remarkably accurate determinations. It is undoubtedly the future method for determining the solar parallax.

But it is above all else in the study of the individual motions of the stars that Doppler's principle has, up to the present time, rendered its greatest service to astronomy. Without its assistance we would know virtually nothing of the motions of the stars and of the interstellar gas that envelops them; we should almost certainly be ignorant of the rotation, and therefore mass, of the Galaxy and of the extragalactic nebulae; and we should lack all observational evidence of the expansion of the Universe.

A few examples of the application of the Doppler effect will make clear its value to astronomy. The English amateur, Huggins, was in 1868 the first astronomer to measure visually the spectral shifts of several stars caused by the Doppler effect. In 1871 the German astronomer Vogel demonstrated the accuracy of Doppler's principle by determining the tiny relative displacement of the solar lines at the east and west limbs of the Sun — the former approaching the Earth and the latter receding, as a consequence of the Sun's axial rotation. The value of the Sun's rotation period derived by Vogel in this way agreed perfectly with that based on visual observations of sunspots. Finally, the American astronomer Young in 1876 provided the conclusive evidence for the validity of Doppler's principle by showing that the telluric lines in the solar spectrum do not partake of this displacement; this is exactly what would be expected, since the Earth's atmosphere is stationary relative to the observer. This behaviour of telluric lines allows them to be distinguished infallibly from the Fraunhofer lines, which are of solar origin — a technique that has been exploited with particular success by the French physicist, Cornu.

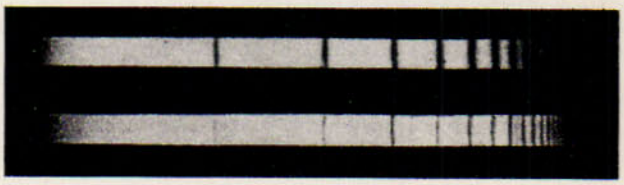


Fig. 798 — COMPARATIVE WIDTHS OF THE LINES OF THE BALMER SERIES OF HYDROGEN. Above, a dwarf, ζ Ursae Majoris, type *d*A2; below, a supergiant, α Cygni, type *c*A2. *(Photos, Chalonge and Barbier, Jungfraujoch; 1935)*

It will readily be understood that the measurement of such microscopic shifts (less than 0.1 Å in the case of the Sun's rotation) was a matter of the greatest practical difficulty so long as they had to be made visually. But since the application of photography they have become perfectly practicable and simple. At the present time the measurement of Doppler shifts is normally made from negatives with the aid of a special instrument — the Hartmann spectrocomparator — which permits the visual juxtaposition of the two spectra that are to be compared: one, for instance, being that of a star whose radial velocity is to be measured, and the other a reference spectrum which is usually that of a laboratory source, or sometimes that of a star whose radial velocity is already known. These standard stars, the publication of whose data is nowadays in the hands of the International

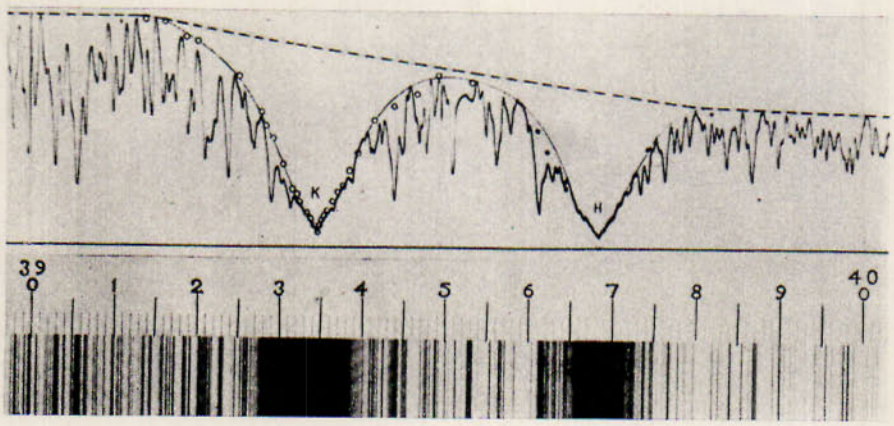


Fig. 799 — THE SOLAR SPECTRUM IN THE REGION OF THE H AND K LINES OF IONISED CALCIUM (3900–4000 Å). Above, trace of the energy curve; below, direct spectrogram from the Rowland *Atlas*. Note the considerable width of the H and K lines, whose wings overlap at about 3950 Å. *(Pettit, Mount Wilson Observatory; 1940)*

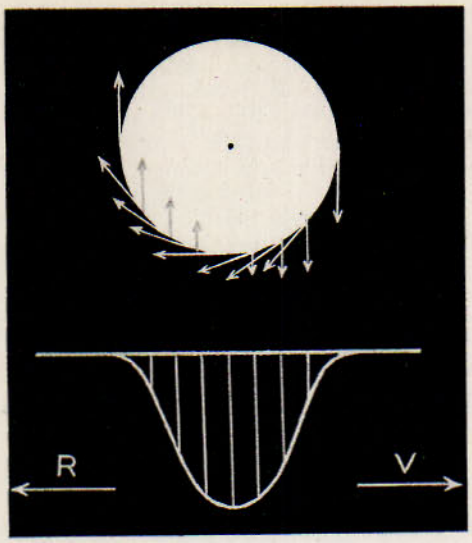


Fig. 800 — PRINCIPLE OF THE WIDENING OF THE SPECTRAL LINES OF STARS IN RAPID ROTATION. To every zone of the star there corresponds a line which is displaced to some extent towards either the red or the violet according to the sign and value of its radial velocity; the sum of all these displacements is a widened line whose profile is of the type illustrated.

Astronomical Union, play an analogous role to the stars of the International Polar Sequence in stellar photometry: they are stars whose radial velocities have been very accurately determined, many times and by many different observers, and have been shown to be constant. The latter qualification is necessitated by the fact that radial velocities of certain stars are variable, even after the effect of the Earth's circumsolar motion has been allowed for. This variation of radial velocity springs from one of two main causes.

The first is analogous to that produced by the Earths' annual revolution about the Sun: the star whose spectrum is being measured is a component of a binary star, revolving about the system's centre of gravity with a period and linear velocity that are determined by the masses of both components and their distance from one another, in obedience to Kepler's laws. Such a star is, in fact, a member of a spectroscopic binary, to which reference has already been made in Chapter 13. The phenomenon is particularly noticeable when the spectra of both components are visible, for the lines then appear alternately single and double, as explained in Fig. 792. In the course of the revolution of the two stars about their common centre of gravity their radial velocities are directed alternately along the observer's line of sight and at right angles to it. In the former case, one component is moving directly towards the observer and the other directly away from him; their spectra are therefore shifted from their mean position — that of the approaching star towards the violet and that of the receding one towards the red; if the two stars are of the same spectral type every line will therefore appear double. In the latter case the lines of the two spectra will coincide, since the radial components of their orbital velocities are nil, and they will therefore appear single and undisplaced (Fig. 792).

Each spectral line will thus oscillate back and forth on either side of its mean position during the course of one revolution, and the relation between the maximum displacements at once gives the relative masses of the two stars. Detailed mathematical analysis of the radial velocity curves obtained by plotting the successive velocities observed throughout a revolution is even capable of revealing the exact values of nearly all the elements of the system as surely as if the two components were individually visible.

Even when one component is much brighter than the other, so that its spectrum alone is visible, the fact that it belongs to a binary system is betrayed by the regular variations of its radial velocity: the lines of its spectrum will oscillate back and forth about their mean position in a period that may be short or long. Even in these circumstances most of the elements of the system can be calculated.

It was therefore believed for a long time that all stars with variable radial velocities were components of binaries. But as soon as the progress of physics allowed the linear dimensions of certain stars to be measured it became apparent that some of the calculated orbits were actually smaller than the stars themselves: this implied that the two stars revolve one inside the other, which is absurd. Now, the stars in question were all variables of a familiar

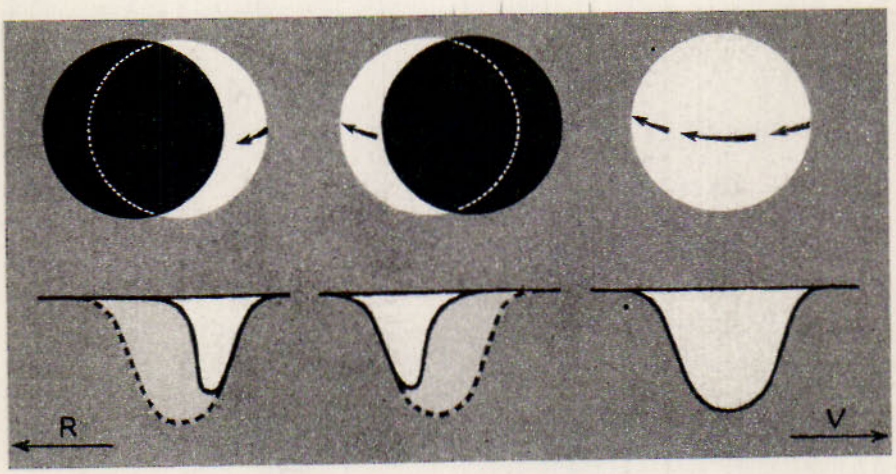


Fig. 801 — DEFORMATION AND DISPLACEMENT OF THE SPECTRAL LINES OF AN ECLIPSING BINARY AS A RESULT OF ITS ROTATION. The profile of a line, widened by the star's rotation (right), is during eclipse successively modified as the line's 'centre of gravity' is shifted first in one direction, then in the other, depending on whether the uneclipsed section of the star is approaching the observer (left) or receding from him (centre).

type — the Cepheids, which have been described in detail in Chapter 13. It was as a result of this discovery that Shapley and Eddington suggested that we are here concerned with pulsating stars, and that the variable radial velocity is a result of this pulsation and of nothing else: the hemisphere of the star that is turned towards the observer is alternately approaching and receding from him. This suggestion was also shown to provide a very satisfactory account of the variations of brightness and colour exhibited by the Cepheids, and it is fundamental to most recent theories of stellar structure (see p. 321).

Other variable stars (e.g. long period variables) present more complex problems, since the displacements of the lines of different elements indicate neither the same value nor the sign of the radial velocity. The interpretation of this extraordinary state of affairs is based on the fact that the stars concerned are all giants — that is to say, stars which are transparent to considerable depths, and which therefore allow the different motions of their component gases to be observed.

This phenomenon plays a particularly important role in the understanding of certain categories of star which we have already described — P Cygni variables and, above all, novae. The spectra of the latter, during their explosive phase, characteristically contain bright and very wide lines, often multiple, which are flanked on their short-wave side by absorption lines of the same element. Later, when the nova has faded considerably, bright nebular lines, produced by the expanding gaseous envelope, are observed to be *doubled* — a phenomenon that is seen even more clearly in the spectra of planetary nebulae. This is easily explained on the assumption that these immense, expanding bubbles of gas are transparent, so that we receive radiation both from the near side (situated between the star and the Earth) which is approaching us, and from the further side (beyond the star) which is receding. The two opposite radial velocities produce a doubling of the lines similar to that which we have already encountered in connection with spectroscopic binaries.

So much by way of explanation of the astronomer's demand for large numbers of stellar radial velocity determinations.

The making of these determinations was, until quite recently, a long and arduous undertaking — each star had to be dealt with individually with a slit spectrograph, which is always a bothersome instrument to use. But within the last few years the French astronomer, C. Fehrenbach (deputy director of the Haute-Provence Observatory), has developed a method of making sufficiently accurate determinations of stellar radial velocities from photographs taken with an objective prism — and each such photograph carries the spectrum of every star in the field photographed.

The procedure consists of photographing the field and then inverting the prism and taking a second photograph, slightly displaced from the first in a direction at right angles to the dispersion of the spectra; the two spectra of a given star are thus reversed end to end. The relative displacements of the lines of the same elements are then measured: this displacement will correspond to double the distance that would separate the images of an ideal line that had undergone no shift during the rotation of the prism.

When it is learnt that the distances measured on the plates amount to only a few hundredths or even thousandths of a millimetre, it will be appreciated that the successful application of the technique, which is simple enough in principle, demands the highest degree of precision in the construction, optical qualities and mounting of the prism. It must, in particular, be absolutely free from distortion.

C. Fehrenbach and A. Couder succeeded in making such a prism — it was in fact a compound prism constructed of two different types of glass whose distortions neutralise one another — which they termed a 'normal field prism'. The line remaining undisplaced by the rotation of the prism corresponds to the $H\gamma$ line of hydrogen when the star's radial velocity is nil.

With prisms in use at the present time it is possible to reach magnitude 10, the mean accuracy being of the order of 2 m.p.s. for a single star. A new prism, whose refracting edge measures 12 inches, is at present under construction, and will in the near future begin to provide a mass of new data in the field of stellar radial velocities.

Fig. 794 is a photograph of the type obtained with an ordinary objective prism. The photograph illustrated in Fig. 795 was obtained with a rotating Fehrenbach prism; each star is represented by two spectra, lying close alongside one another, but reversed end to end.

These few examples will suffice to show how powerful a tool spectroscopic analysis is in the hands of the astronomer, but they are far from exhausting its potentialities. By its means we can detect and investigate much more delicate effects, and it is to these that we must now turn our attention.

Other spectroscopic effects; line widths in astronomical spectra. — The versatility of the spectroscope as a research instrument depends on the fact that the atoms whose radiations it analyses are themselves extremely complex and delicate structures, sensitive to the least external perturbation. Some of these perturbing influences, chosen from among the many that are relevant to astronomy, are described below.

The most important is that produced by a magnetic field; this was discovered in 1896 by the Dutch physicist Zeeman. Its essential features are as follows. It is found that when the atoms of a radiating or absorbing vapour are exposed to a magnetic field, the lines of its spectrum are split into two, three or more close-set components which are arranged symmetrically on either side of the original line; moreover the radiations producing these lines are polarised differently from one another. We cannot enter here into the detailed theory of the Zeeman effect, for this would lead us too far out of our way; but, very briefly, the spectroscopic phenomenon results from the fact that the energy of an atom in a magnetic field is different from that of the same atom when not situated in such a field, and depends upon the orientation of the electron orbits relative to the field. It has been shown that a limited number of well defined orientations are alone possible (spatial quantification); as a result, a magnetic regrouping of the atom's energy levels takes place, and this in turn is responsible for the splitting of the spectral lines. Finally, both theory and experiment show that in fields whose intensities do not exceed a few thousand gauss the separation of the lines is proportional to the field strength. This valuable relationship allows the intensity of magnetic fields to be determined simply by measuring the separation of the components of the spectral lines.

The first extra-terrestrial magnetic field to be detected by this method was that of the sunspots — by Hale, at Mount Wilson, in 1908. Hale noticed that numerous lines in high-dispersion spectra of spots were doubled, and confirmed that the polarisation of their components was that predicted by theory; the separations of the components indicated field strengths of the order of 4,000 gauss. We have already explained that these fields are attributed to the vortical motions of the electrons in sunspots (Chapter 10).

The separation of the components of a spectral line by a magnetic field is at the best of times very small, and it can only be detected (even in high-dispersion spectra) when the field is intense: if either the instrumental dispersion or the field strength is inadequate, the line will merely appear to be widened slightly. The interpretation of the effect is then highly intricate, for numerous other agencies cause a widening of the lines.

An examination of the spectra reproduced in this book will suffice to demonstrate that spectral lines differ from one another in the matter of their width; some are very fine, others occasionally very wide; but even the finest always have a measurable width. This width is the effect of many causes, both instrumental and natural.

The instrumental causes (slit width, diffraction and diffusion) can be investigated in the laboratory: their magnitude can be determined and their effect allowed for in the final results, so that we can ignore them throughout the following account, and assume that we are using a theoretically perfect instrument.

But apart from this, the lines in astronomical spectra are widened for reasons that are inherent in the nature of things. These may be summarised as follows:

(1) Spectral lines have 'natural' widths resulting from the fact that atoms, like radio stations, do not emit radiations of a single wavelength but a restricted gamut of wavelengths which are concentrated about a mean value. This is termed 'natural damping'. The natural widths of spectral lines are of the order of 0.002 Å.

(2) The atoms of a radiating or absorbing gas have rapid random motions whose mean velocity is a function of temperature — the so-called thermal agitation of the medium. Thus the wavelengths of the radiations emitted or absorbed by these atoms are to a greater or smaller extent increased or reduced by the Doppler effect, according to the value of the radial component of their motion at the instant of emission or absorption; this component will vary from one atom to the next, and will be continuously changed by collisions. The aggregate effect in a very large number of atoms will be a widening of the lines which may amount to several 0.01 Å at stellar temperatures.

(3) Atomic collisions will have a further effect: it can be shown that if a collision occurs while one of the two atoms is actually in the process of emitting (or absorbing) radiation, the sudden interruption of the process will result in a slight modification of the wavelength of the observed radiation. This is the so-called 'broadening by collision'. It becomes particularly noticeable when the density (and therefore pressure) of a gas is relatively high, for then, obviously, collisions will occur much more frequently. Hence this effect is weak in the spectra of giant stars, whose atmospheres are highly rarefied, but becomes noticeable in those of the much denser dwarfs. This difference provides us with a spectroscopic criterion for distinguishing between giants and dwarfs (p. 342).

(4) This distinction is further, and very strongly, accentuated by another optical phenomenon — the Stark effect — which consists of the splitting of the spectral lines under the influence of an electrical field; it is very similar to the Zeeman effect *vis-à-vis* magnetic fields. Large-scale electrical fields do not (unlike the magnegic fields associated with sunspots) occur in stellar atmospheres, but we know that at the high temperatures obtaining in these atmospheres a large proportion of the atoms are ionised; these are no longer electrically neutral, and therefore create around themselves small electrical fields which are capable of affecting other atoms which are continuously passing in random directions through their vicinity. Thus it happens that the radiations of these atoms are modified by the Stark effect. The combined

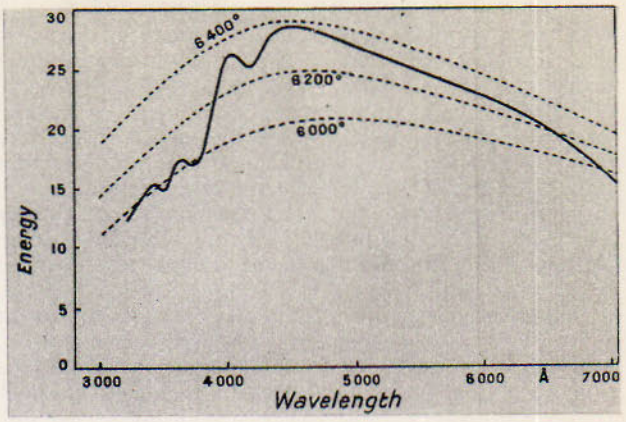


Fig. 802 — ENERGY DISTRIBUTION IN THE CONTINUOUS SPECTRUM OF THE CENTRE OF THE SUN'S DISC, compared with the energy curves of a black body at different temperatures. *(Pettit, Mount Wilson Observatory; 1940)*

effect of all these individual atomic modifications is simply a widening of the line, and, again, it will obviously be inconsiderable if the mean separation of the atoms is large, i.e. if the pressure is low, as it is in the atmospheres of the giants. Contrarily, it will become very marked — exceeding several angstroms — when the atoms are closely packed, as is the case in the high-pressure dwarfs. Such is the origin of the characteristic differences of appearance of the lines in giant and dwarf spectra, already referred to, which are the basis of the powerful spectroscopic method of determining stellar parallaxes (Chapters 12, 14).

(5) There is still another cause of line widening. The spectra of certain stars on the dwarf branch (particularly those of types B and A, such as Algol and Altair) contain wide, diffuse lines, while those of other stars of the same type and general characteristics present narrow and well-defined lines. Obviously, in this case, the Stark effect cannot be invoked. The solution was provided by O. Struve in about 1925, and concerns the axial rotations of the stars. It is an obviously legitimate assumption that the stars, like the Sun, are rotating, and that their rotational axes are randomly oriented in space: some will be pointed towards the Earth, others at right angles to the line of sight, and others again will be found at very intermediate inclination, the distribution being completely haphazard. Let us consider the case of a star whose axis is more or less perpendicular to the line of sight. It will be clear (Fig. 800) that one limb will be approaching the observer while the other will be receding — a state of affairs that we have already encountered in the case of the Sun; different points on its visible hemisphere will be approaching and receding at every velocity between these two extremes. A Doppler effect which varies across the disc from limb to limb will therefore be produced, and since every point on the disc contributes to the formation of the observed spectrum the total effect will be a widening of the lines in accordance with a law whose effects are easily calculable. Comparison of the calculated line profiles* with those in fact observed has shown that this explanation is valid, and in this way some very rapid axial rotations have been detected — these are not infrequently of the order of 60 m.p.s. at the equator, and in some cases even approach 200 m.p.s. (cf. the Sun's equatorial rotation of a little over 1 m.p.s.).

These unexpectedly high values have, however, been

* The intensity distribution across the width of a line.

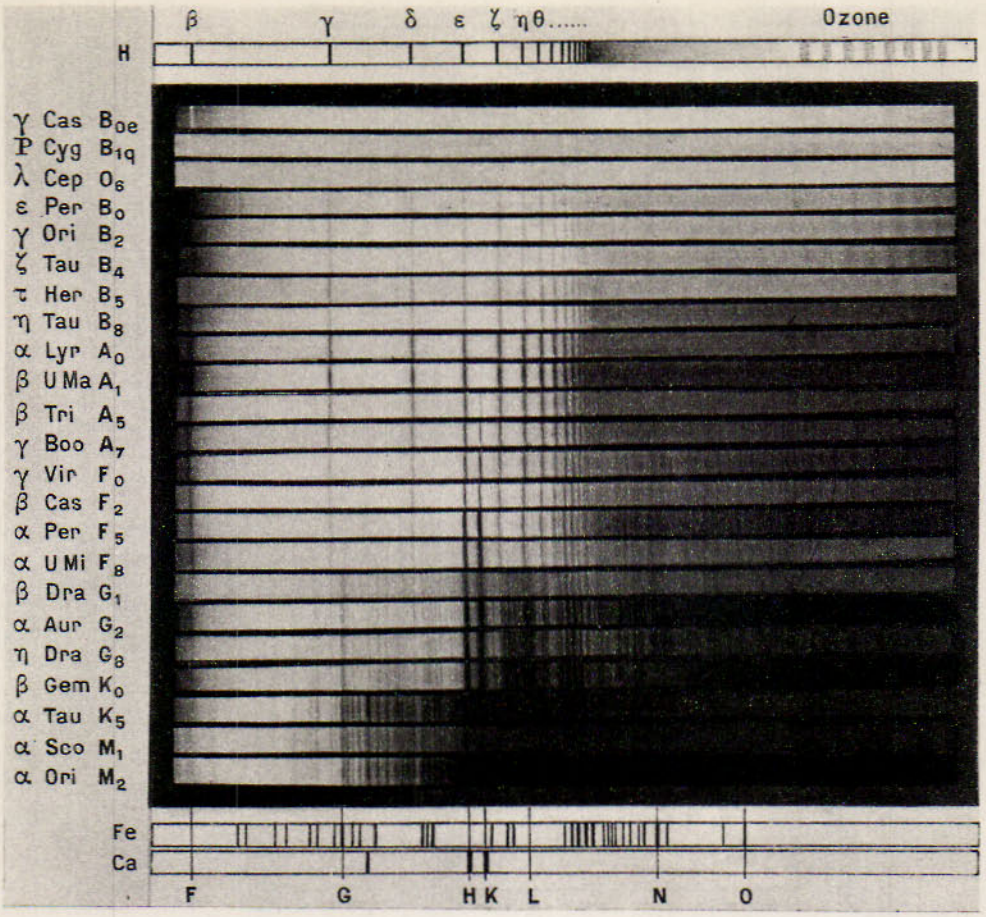


Fig. 803 — SERIES OF STELLAR SPECTRA IN THE ULTRA-VIOLET AND BLUE-VIOLET REGION (3100 to 4900 Å), demonstrating the progressive evolution of the lines of various elements from one end of the spectroscopic classification to the other. Compare with Fig. 804. The diffuse bands which appear (mainly) in the extreme ultra-violet (towards the right) of the B type spectra are due to ozone in the Earth's atmosphere. *(Photos, D. Barbier and D. Chalonge, Jungfraujoch; 1935–8)*

confirmed by independent methods. Only one of these can be described here, and this has to be mentioned because of its ingenuity. The idea was first suggested by the American astronomer F. Schlesinger half a century ago, but it was not successfully applied until 1924, when E. Rossiter and D. B. McLaughlin showed it to be practicable. Consider an eclipsing binary such as Algol, one of whose components is very much brighter than the other; it will appear as a spectroscopic binary, only one of whose two spectra is visible. There will come a moment during the course of the eclipse (which, for the sake of simplicity, we shall suppose to be central) when the 'dark' component hides half the disc of the 'bright' one. At this moment the greater part of the light we receive from the latter will have been emitted by regions that are all receding from us or all approaching us. In the former case the spectral lines will suffer a displacement towards the red; in the latter, towards the violet. And in either case, after mid-eclipse has been passed, the opposite effect will begin to obtrude itself. Under such circumstances it should be possible, by sufficiently detailed analysis of the spectrum at different phases of the eclipse, to determine the velocity of the 'bright' star's axial rotation. This was actually done, and the velocities so obtained were found to be in excellent agreement with those derived from line-widths measured outside eclipse; very satisfactory confirmation of the validity of the line-width method was thus provided.

The reader who has followed the chain of reasoning which permits such accurate and detailed conclusions to be reached will be able to judge for himself of the astonishing versatility and precision of modern applications of spectroscopic analysis.

But its resources do not end here — they are, indeeed, almost limitless. Thanks to increasingly delicate and perfected techniques (which we cannot, for obvious reasons, describe in detail) spectroscopic analysis is every day providing the astronomer with fresh information about the Universe. We must content ourselves with citing two final examples, which are particularly noteworthy.

The first of these concerns the Zeeman effect once again, and the evidence that it has provided for the existence of a general magnetic field in the Sun and the stars. Following his discovery of intense and localised magnetic fields in the sunspots, Hale suggested that the Sun (considered as a whole, and quite apart from the spots) might be enveloped in a weaker, general magnetic field, analogous to that of the Earth. In 1913, with the Mount Wilson 150-foot solar tower, he succeeded in detecting certain minute line displacements (of the order of 0.001 Å) in very wide-dispersion solar spectra, which could be attributed to such a field and which indicated an equatorial field strength not exceeding 50 gauss. In principle, the method consisted in the alternate suppression of the two oppositely polarised but not visually

separated components of the widened line, by means of a suitable polariser and analyser; in this way the centre of the widened line will be displaced towards on side of the component which has been retained. These extraordinarily delicate measurements remained for a long time unique, and it was not until 1945 that the German astronomer G. Thiessen repeated and completely vindicated them. There is therefore no doubt about the existence of the Sun's general magnetic field.

The question naturally arose whether the stars also are not surrounded by magnetic fields, particularly as certain physical theories suggest that any rotating body should generate a magnetic field whose strength is proportional to its speed of rotation — and, as we have seen, B and A type stars are known to be rotating with a mean equatorial velocity of about 45 m.p.s. If such a star were so oriented in space that its axis were pointing towards the Earth there would be no Doppler effect to widen its lines, since the rotational motion of every point on its surface would occur in a plane at right angles to the line of sight, and would therefore have no radial component. At the same time, one of its two polar regions would be directly facing the Earth, and these are the very regions where the star's magnetic field, if any, would be most intense and the splitting of its lines by the Zeeman effect most noticeable. Choosing a type A star (78 Virginis) whose narrow lines suggested a suitable orientation of the rotational axis, the American astronomer H. W. Babcock was in 1946 able to detect (on wide-dispersion spectrograms obtained with the Mount Wilson 100-inch, and using a technique similar to Hale's) the existence of a 1,500-gauss magnetic field surrounding the star. This value agreed well with theoretical predictions.

The red giant, β Pegasi, on the other hand, presents no detectable effect. This again was predicted by theory, for stars of this type always have very fine spectral lines, and must therefore have correspondingly slow rotations.

The second revelation of spectroscopic analysis which must be mentioned — it is even more important, in view of its consequences, than the foregoing — is the so-called Einstein effect. This concerns the influence that gravitational fields exert on the period of the vibration of different radiations, as predicted by the theory of general relativity. One conclusion derived from this theory is that the measurement of time will actually be different in an intense gravitational field and in a weak one such as the Earth's natural phenomena, such as the oscillations of a pendulum, the vibrations of an atom, etc., will in the eyes of an external observer be slowed down in an intense gravitational field; the effect of this on emitted radiation will be an increase of period or (the same thing) a decrease of frequency, corresponding to an increase of wavelength and therefore to a shift of the spectrum towards the red, resembling that produced by a positive radial velocity.

The earliest attempts to discover observational evidence of this effect were concentrated on the Sun, whose surface gravity is 28 times that on the Earth. But it became evident that the tiny spectral shift produced by a gravitational field of this strength would be effectively masked by the Doppler shifts produced by the motions of the gases in the Sun's atmosphere. Happily, however, there are stars whose surface gravitational fields are incomparably more intense than the Sun's — the white dwarfs. In Chapter 14 we considered the characteristics of Sirius B, as being representative of this class of stars: although its mass is equal to that of the Sun, its radius hardly exceeds 2 or 3 times that of the Earth; the atoms at its surface are therefore subjected to a gravitational field whose intensity is about 30,000 times that of the Earth's, and theory predicts that the displacement of the lines in the blue-violet region of its spectrum should amount to 0.3 Å; if this were a Doppler shift it would indicate a velocity of about 12 m.p.s. — easily detectable and measurable. In the case of Sirius B, however, the observation is rendered extremely difficult by reason of the proximity of its brilliant neighbour, whose light (diffused both in the terrestrial atmosphere and in the telescope) tends to swamp that of the dwarf in the spectroscope slit, and also because the spectral lines of white dwarfs in general are strongly widened by pressure (the Stark effect, depending on atomic collisions). However, this extremely delicate observation was made successfully by Adams in 1925, with the aid of the Mount Wilson 100-inch, and he was able to confirm that when allowance is made for Sirius's radial velocity and the orbital motion of its companion (both of which are well determined) there is a residual red shift which agrees exactly with that predicted by Einstein. Adam's observation demonstrated simultaneously the validity of Einstein's equations and the reality (by no means well established at that time) of hyperdense dwarf stars. Since then, the Einstein effect has been detected in the spectra of other white dwarfs, and it has even been employed to determine the masses of these stars when they are not members of binary systems.

But we still have not exhausted the inventory of the revelations provided by light itself, once astronomers learnt to interpret spectra. Indeed, we have so far only considered the positions of the spectral lines and the factors that may modify these. Line-width, however, is only one characteristic of a spectrum: lines (whether emission or absorption) differ among themselves as regards their intensity; furthermore, the intensity distribution in the continuous background also varies in a recognisable way according to the conditions under which the radiation producing it is emitted. We must now inquire, though summarily, what further information can be extracted from the study of spectra.

Energy distribution in continuous spectra line intensities. — We have already learnt in Chapter 14 that the colour of a star is an indication of its surface temperature, and that the latter can be derived from measurements of the former. This subject is worth looking at in a little more detail.

According to Kirchhoff's radiation laws any body is capable of absorbing the same radiations that it emits; the precise relation that Kirchhoff established between these two processes is such that the emission and absorption capacities of a body are exactly equal. It follows from this that a body capable of absorbing *all* the radiations it receives — i.e. an absolutely black body — will also be a perfect radiator, in the sense that it will emit, at any given temperature, the theoretical maximum of radiation at every wavelength. This hypothetical perfect radiator is therefore termed a 'black body' — which in the circumstances must appear rather a strange name unless the reason for it is understood.

Black-body radiation is obtained in the laboratory by heating to the required temperature a fireproof ceramic tube in which a small hole has been made: a ray of light entering the tube through this hole will have virtually no chance of escaping again, but will ultimately be absorbed after undergoing a series of successive reflections from the inner wall; and conversely, the radiation which does emerge from the interior of the incandescent tube will be that of a black body. The inspection window of an industrial furnace is to all intents and purposes a black-body radiator.

Until the end of the nineteenth century no theoretical

explanation of the energy distribution in an experimentally produced black-body spectrum had been discovered. In attempting to do so, the German physicist Max Planck in 1900 made the then revolutionary suggestion that atoms do not emit radiation in the form of a continuous flow but discontinuously in the form of discrete packets, each of which is composed of a whole number of small and indivisible units of energy. This was the quantum theory, which all recent developments in physics have served only to extend and refine.

Starting from this assumption, Planck was able to derive a theoretical expression for the precise distribution of energy throughout the spectrum of a black-body radiator as a function of its temperature. This distribution, for three different temperatures, is shown in Fig. 802 (dotted curves).

Granted all this, it is easy to understand the principle and results of the spectrophotometric measurement of stellar temperatures. We saw in Chapter 15 that the continuous component of a star's spectrum (i.e. the bright background on which the absorption lines are superimposed) is produced by radiation from the zone, of finite depth, which constitutes the star's photosphere and throughout which the density increases from the exterior inward. A light ray of any wavelength which penetrates this layer from the exterior will sooner or later be totally and irremediably absorbed; the stellar photosphere may therefore be regarded as constituting a black body. In principle, therefore, it is only necessary to compare the energy distribution of its continuous spectrum with those of theoretical black bodies at different standard temperatures in order to discover its temperature. This temperature will simply be that of the black body whose energy distribution curve most closely matches its own (Fig. 802)

If this curve coincides with one of Planck's theoretical curves — at least over a material part of its spectral range — the so-called 'colour temperature' of the star can be derived. This is the temperature of the black body whose colour is the same as the star's — the observed colour of a source being determined, of course, by the proportionate intensities of the different wavelengths that comprise its radiation, which is what is represented by its spectral energy curve.

The temperature derived in this way is not necessarily, however, the true temperature of the star's radiating layers: (1) because, as we have already learnt, the observed radiation is emitted at different levels within the star, whose temperatures are themselves different; (2) because the distribution of energy in a continuous spectrum is often difficult to plot accurately if the spectrum is cluttered up with innumerable absorption lines — as is the case with stellar spectra of type later than A; with stars of types K or M, or later, it becomes in the last degree uncertain, if not quite impossible to evaluate, since the continuous photospheric spectra of these stars are literally mutilated by the intense molecular bands superimposed on them by the stars' atmospheres; (3) most important of all, because even in the case of early-type stars it is by no means certain that the continuous spectrum as observed is a true reflection of the photospheric radiation, since the spectrophotometric trace can be materially affected by 'continuous absorption' extending over a wide wavelength range, caused by unsuspected components of the stellar atmosphere. It has already been mentioned (Chapter 15) that within the past 20 years or so at least one such component (the negative hydrogen ion) has been recognised as exerting absorption throughout the whole visible and infra-red regions of the majority of stellar spectra other than those of type B stars. It had been recognised at an even earlier date that the continuous spectra of B, A and F type stars are modified by a hydrogen *atom* absorption (or emission) which extends beyond the limit of the Balmer series in the near ultra-violet.*

The upshot of all this is that the observed colour of a star has no more than a tenuous connection with its actual temperature, and the derived values do in fact

* The theory of atomic spectra indicates that this continuous absorption is connected with the photo-ionisation of the atom — i. e. its ionisation through the agency of the photoelectric effect. The mechanism of this type of ionisation is as follows. When a photon of larger energy than that binding an electron to the nucleus is absorbed by the atom, it tears the electron out of the atomic structure and ejects it with a kinetic energy which is proportional to the excess of its energy over that just neccessary to produce ionisation; this is the energy corresponding to the limit of the converging series of lines. In other words the electron is expelled from the atom at a velocity which is inversely proportion to the wavelength of the incident radiation, and since the velocity of a *free* electron is not quantified (i. e. may assume any value from zero upward) the result is a *continuous* absorption spectrum.
Conversely, the capture by ions of free electrons moving at widely differing velocities gives rise to a continuous emission spectrum.

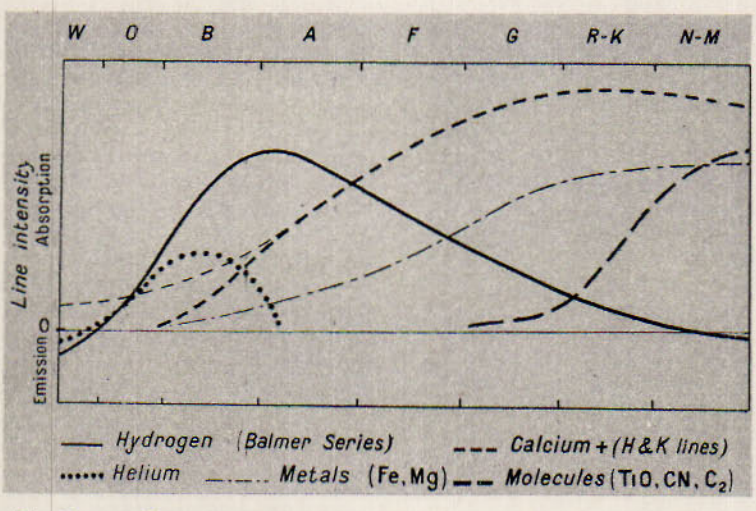


Fig. 804 — GENERAL APPEARANCE OF THE INTENSITY-VARIATION OF THE LINES AND BANDS OF THE PRINCIPAL ELEMENTS AND COMPOUNDS ALONG THE SPECTROSCOPIC SERIES *(Rufus, Observatory of Michigan, Ann Arbor; 1923)*. The extension, in the light broken line, of the curve for ionised calcium represents the observed intensity due to interstellar calcium. Note that in types W and M emission lines of hydrogen make an appearance. The ordinates scale is arbitrary.

differ rather widely according as to whether the measurements are made in the ultra-violet (3100–3700Å) or the blue-violet (3700–4600Å) region of the spectrum. Here, for example, are the colour temperatures of B, A and F type stars, as derived from measurements made in these two regions by the French astronomers D. Barbier and D. Chalonge (1935–8):

REGION OF SPECTRUM	SPECTROSCOPIC TYPE		
	FO	AO	BO
Ultra-violet	20,000°	10,500°	8,300°
Blue-violet	28,000°	16,500°	8,600°

However, there are several other methods* of determining stellar temperatures, based on a study of line intensities. The theory of these, very briefly, is as follows:

We have already seen that the production of the successive spectra of an element — first neutral, then progressively ionised — is controlled by temperature. Consider any element (calcium, for instance) subjected to a temperature of some 5,000° in the reversing layer of a star such as the Sun. At this temperature some of its atoms will still be neutral, and will therefore emit (or rather, in this case, absorb) radiation corresponding to the lines of its neutral spectrum, Ca I; these include the line, familiar to astronomers, at 4227Å. But the majority of its atoms will already be ionised, having experienced collisions of more than average violence, and these will produce the emission (or absorption) lines of the Ca+ ion, i.e. the Ca II spectrum; among these are the widest and most intense of all the solar lines, Fraunhofer's H and K lines at the extreme long-wave end of the visible spectrum (3968 Å and 3933 Å).

It will be appreciated that the relative strengths of the lines of these two spectra will depend on the temperature: raising the temperature will favour ionisation and therefore increase the intensity of the Ca II spectrum, while weakening the Ca I spectrum, and vice versa.

We saw in Chapter 14 that all the elements undergo analogous spectroscopic metamorphoses; but the temperatures at which these changes occur, and the positions of maximum intensity in the successive spectra, vary from element to element, since they depend upon the ionisation potentials and these, in turn, are determined by the structure of the atoms (Fig. 804). From these differences — with the help of a formula established theoretically by the Indian astrophysicst M. N. Saha in 1920 (which gives the relative proportion of ionised to neutral atoms of a given element as a function of its ionisation potential and of temperature) — it is possible to deduce the precise temperatures to which the atoms are being subjected on stars of the various spectral types.

It will readily by seen that by comparing Saha's curves (showing the variation with temperature of the relative intensity of carefully selected lines of different elements at different ionisation levels) with the observed variations along the spectral series, one could, in theory obtain the required temperature with considerable accuracy. Such temperatures are termed 'ionisation temperatures' — a choice of name which is both appropriate and prudent, for here once again the matter is more complex in practice than one might suppose from a description of the elementary theory.**

It must be mentioned, however, that Saha's equation includes an important term which takes into account the electron pressure in the gas. If the atoms in a star's atmosphere are more or less completely ionised, that atmosphere will necessarily contain a large number of free electrons that have been torn out of the ionised atoms; and their number will increase with the proportion of ionised to neutral atoms, i.e. with increasing temperature. But owing to the high temperature these electrons will have violent random motions and will therefore constantly be involved in collisions with atoms and ions. The latter will immediately recapture the electrons which they had lost as a result of earlier collisions. Such collisions leading to the recombination of ions and electrons, will thus tend to counteract the effect of high temperature.***

It is obvious that this contrary effect will be maximal when the electrons are most numerous, and when, therefore, the pressure that they exert by their collisions is high. It is for this reason that a term for electron pressure had to be incorporated in Saha's ionisation formula.

To conclude this survey we shall mention one very remarkable consequence of electron pressure, since it leads straight to the spectroscopic determination of absolute magnitudes, and therefore of parallaxes. The importance of this in mapping the Sun's neighbourhood of the Galaxy has already been emphasised in Chapters 12 and 14.

Consider the case of two stars of the same type (G, let us say) and roughly the same temperature, but differing from one another in being, one, a giant, and the other a dwarf. We know that the pressure in the distended atmosphere of the giant is almost negligible compared with that in the relatively enormously dense reversing layer of the dwarf. There will be a corresponding disparity between the electron pressures in the two environments: ions in the giant's atmosphere will have a relatively poor chance of encountering a free electron to replace the one they have lost; in the atmosphere of the dwarf, on the other hand, collisions will be much more frequent and the opportunities for recombination correspondingly greater. As a result, other things being equal, ionised atoms will be relatively more numerous in the giant than in the dwarf. As a further result — and it is here that the crux of the matter lies — the lines of ionised atoms will be more intense (compared with those of neutral atoms) in the giant spectrum than in the dwarf. In theory, for example, the Ca I line of 4227Å will be more intense, compared with the H and K lines of Ca II, in a dwarf such as the Sun than in Capella A, a giant of the same type. In practice, however, the situation is once again somewhat more complex than in theory, for giants and dwarfs of the same spectral type do not in fact have exactly the same surface temperatures (Chapter 14), that of a giant being slightly lower than that of a dwarf. But the validity of the principle remains unaffected, and observation has proved that there is a systematic difference between the relative intensities of suitably chosen lines

* It is nowadays possible to derive the effective temperature of a star from its observed colour temperature, thanks to our improved knowledge of stellar atmospheres. We can do no more than mention, here, such other definitions of temperature (derived from Planck's equation) as luminosity temperature and radiation temperature.

** It may be mentioned that an analogous formula may be applied, not to the different ionisation states of a single atom (i. e. to its successive spectra), but to the different excitation states of an ion or atom, i. e. to its various energy levels and their corresponding series of lines; the intensities of these also depend upon temperature. This method yields the so-called excitation temperature, which should in theory agree (or at any rate, very nearly agree) with the ionisation temperature. It is also possible to calculate the relative intensities of the lines in the band spectra of molecules, and thence derive the 'rotation temperature' and 'vibration temperature' — a technique which is particularly useful for deriving the temperatures of the coolest stars and even of the atmospheres of planets.

*** It might appear at first sight that these two effects, both of which are the result of thermal agitation, should on the average balance one another. But a moment's reflection will show that this cannot be so, since we are here dealing with phenomena which are subject to the laws of chance. Furthermore, there is an additional ionising agency (p. 495) which becomes very important in hot stars — photo-ionisation, or ionisation by the photoelectric effect — which is promoted by short-wave radiation and which becomes predominant in such bodies as the planetary nebulae (Chapter 16).

in the spectra of giants and of dwarfs belonging to each spectroscopic type; this regular variation can even be traced, as a function of absolute magnitude, among intermediate stars of types B, A, F and sometimes G. The difference has of course to be calibrated by means of stars whose absolute magnitudes have already been determined by some independent method (Chapter 12), but once this has been done a simple comparison of the relative intensities of two of the standard lines will immediately yield the absolute magnitude and therefore the distance of the star.

Even now we are far from having exhausted the long list of discoveries that astronomers and physicists have extracted from the faint but so eloquent light that filters down to us from the stars. And with each passing year the analysis of this light, both in theory and practice, becomes increasingly sensitive and far-reaching.

We have said nothing of the remarkable achievements of quantitative chemical analysis as applied to the stars — based on 'curves of growth' which in turn are the offspring of the study of line intensities — during the past quarter century; an adequate exposition of the rather intricate technique involved would take us too far afield. Nor have we been able to say anything of the subtle logic that has enabled astronomers to deduce, from the mere presence of 'forbidden' lines of oxygen ('nebulium') in planetary nebulae, the extraordinarily high temperature of the often almost invisible central star whose intense ultra-violet radiation excites the nebula to fluorescence. We have done no more than mention the curious spectra of the stars with permanently distended atmospheres (Wolf-Rayet, P Cygni stars, novae), the emission spectra of such stars as γ Cassiopeiae and Mira Ceti, and many others which it is impossible to bring within the framework of this book.

Nevertheless, it is hoped that this final chapter will have given the reader a clearer understanding of the methods and problems of modern physical astronomy.

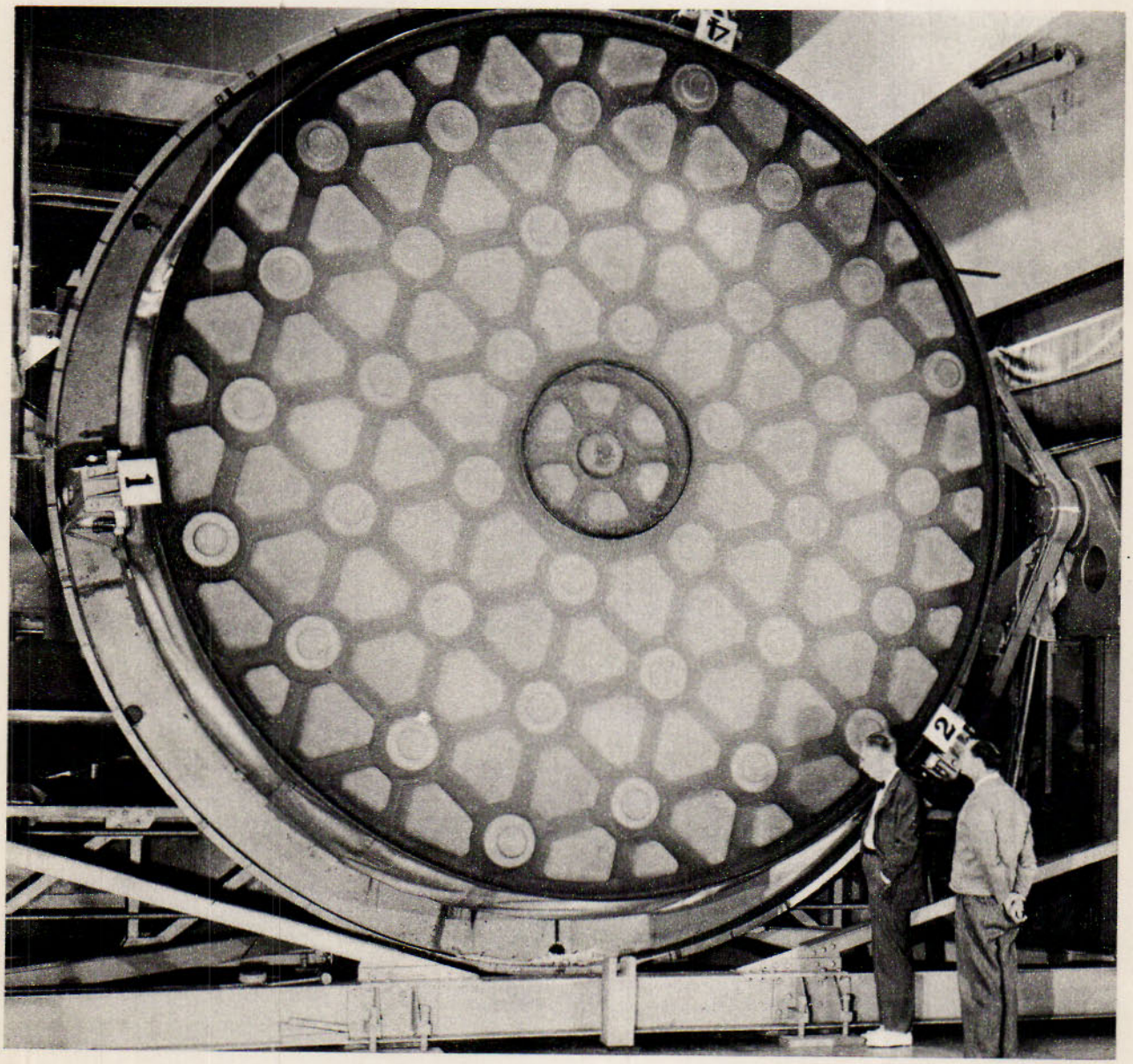


Fig. 805 — THE 200-INCH MIRROR OF THE MOUNT PALOMAR TELESCOPE. Through the transparent body of the mirror can be seen the honeycomb pattern cut out of the back to lighten it; even so, it weighs 15 tons.
(Photo, California Institute of Technology)

'It is indeed a feeble light that reaches us from the starry sky. But what would human thought have achieved if we could not see the stars . . . ?

JEAN PERRIN

CHAPTER 21

THE PHILOSOPHICAL, SCIENTIFIC AND TECHNICAL VALUE OF ASTRONOMY

by G. de Vaucouleurs

Having reached the end of this survey of our present-day astronomical knowledge, and of the principal means by which it has been acquired, the reader (even though his curiosity has been satisfied) may still have one further question to ask: What is the good of astronomy? And because this question is both perfectly legitimate and frequently asked, an answer should be given.

Why should we concern ourselves with these faint and remote points of light, themselves so utterly unconcerned with us and with our daily lives? Why spend so much trouble and money on satisfying what is, at first sight nothing more than useless curiosity?

We may note, to start with, that the layman who a few decades ago was prone to ask the same question regarding physics, is no longer inclined to do so today: the aeroplane, the radio, the cinema . . . and the atom bomb, have adequately widened the horizons of all whose minds are not incorrigibly hidebound.

The practical applications of astronomy were among its earliest achievements, but these are relatively few in number and not particularly spectacular. They may therefore easily be under-estimated, though they are none the less real and worthy of recognition.

The most ancient was the accurate measurement of time, which still plays a fundamental role in our everyday life; the divisions of the calendar, as well as the exact hour, are determined by meridian observations which have been accumulated for several centuries and which are still regularly undertaken at specialised observatories in all parts of the world. It was at the Paris Observatory, headquarters of the International Time Bureau, that the speaking clock was first put into operation, some twenty-five years ago, and its voice relayed daily to the millions of radio listeners. Similarly, it was in response to the clockmaking industry that the observatories at Besançon in France, at Neuchâtel in Switzerland, and Greenwich, in England, have for many years specialised in the regulation of chronometers and watches.

A second practical application of astronomical observations — connected with the first, and no less important — is the determination of longitude and latitude, which is the basis of geodesy, cartography and all forms of maritime and aerial navigation. The title of the annual French collection of fundamental astronomical ephemerides, whose publication has continued without interruption since its inception in 1679, is significant of this: *La connaissance des Temps et des Mouvements célestes à l'usage des astronomes et des navigateurs* ('Knowledge of celestial times and motions, for the use of astronomers and navigators')*.

Yes, it may be said, all this no doubt justifies classical mathematical and positional astronomy, but what of modern astrophysics? What practical use has that? The answer for those who think that science has no value apart from its practical applications is simple: Even astrophysics has a direct relevance to that most 'practical' of human activities — war.

It is not generally known that during the last World War the German military authorities installed a network of solar observatories right across Europe, whose function was to supplement the permanent observatories in providing a complete record of solar activity: for events on the Sun, as we have already learnt, affect the terrestrial ionosphere and hence long-distance radio communication, to a very marked extent — communication with submarines at sea, for example.

In the same way, the jamming of radar sets by sunspots was kept a dead secret by the British authorities during the war, for fear that the Luftwaffe would take advantage of it to launch air-raids of which no warning would be given, the approach of the bombers being masked by the electromagnetic radiation from some large sunspot.

For this reason the physical study of the Sun has, since the war, been pushed forward in every country — and has, alas, already fallen under the shadow of military 'security'.

But fortunately for astronomers — who in this respect are better off than the atomic physicists — the greater part of their science is still innocent of practical and lethal application; nevertheless it provides invaluable support and corroboration to other branches of science. And this, in the eyes of the more enlightened, is the real justification and supreme value of astronomy.

The reason for this is, quite simply, that science is 'one and indivisible'. It is only the complexity of the external world, and the inadequacy of our understanding of it, that forces us to split up the study of the Universe into a series of arbitrary branches — Mathematics, Astronomy, Physics, Chemistry, Biology and the rest. In fact, an advance in any one of these will have more or less direct repercussions in all the others. A few examples will show how true this is in the case of astronomy.

Firstly, astronomy has always been the ideal field for the application of mathematics, and has again and again provided the mathematician with the most perfect practical illustrations of his concepts; not infrequently, indeed, it has been astronomy that gave birth to these concepts. To the naked eye, the Sun and the full Moon are just circular discs. Classical mechanics, which owes its origin to Galileo, found its most complete expression in the celestial mechanics — developed by Newton and a host of famous mathematicians after him; and the study of the stars and the other heavenly bodies still, after more than two centuries, provides them with new problems which demand the invention of new methods of calculation. To take a single example, the famous 'three body problem'

* Similarly with the British *Nautical Almanac and Astronomical Ephemeris*, which has been issued every year since 1767 by Greenwich Observatory.

still remains a challenge to the sagacity of the specialist in celestial mechanics.

Again, in the borderland between mathematics and physics it was an astronomical experiment (in intention, at least) which provided the basis of Einstein's theory of relativity — the famous Michelson-Morley experiment. It was also in the realm of astronomy that this theory found most of its early verifications: the displacement of Mercury's perihelion, the deviation of light rays passing close to the totally eclipsed Sun, and the 'Einstein shift' of the lines in the spectrum of the companion of Sirius.

But if physics has been a rejuvenating stimulus to astronomy, astronomy for its part has not been ungrateful: it offers the physicist those extreme conditions of volume, temperature, and pressure which are unattainable in the laboratory: the millions of degrees and the hundreds of thousands of atmospheres encountered in stellar interiors, as well as the fantastic rarefactions of the interstellar medium, of the nebulae, and of the solar corona — all of which give rise to unforeseen spectra whose interpretation has brought about new advances in our understanding of the atomic bases of spectroscopy.

Astronomy, again, has revealed to physicist the existence of hyperdense degenerate matter in the cores of the white dwarfs, whose theoretical investigation has stimulated the most recent refinements of wave mechanics.

Finally, it is Astronomy that has provided physicists with the most perfect and awe-inspiring examples of natural atomic piles — the stars.

As one passes down the normal classification of the sciences, one finds that their connection with astronomy become increasingly slender and less obvious. Nevertheless, it is always possible to uncover such connections, and these are none the less important for being hidden.

To chemistry, astronomy has contributed at least one new element, helium, which was discovered in the solar prominences twenty-five years before it was isolated in a terrestrial laboratory. Astronomy has likewise demonstrated the occurrence of free ions, or incomplete molecules, in space; though these are unstable on Earth, they were none the less predicted by theory. More recently, the astronomer has detected in certain stars a rich abundance of isotopes which are very rare on Earth.

But the greatest service that astronomy has rendered to the chemists (and equally to the physicists) is its assurance of the universal validity of science — of the occurrence throughout the whole Universe of the same atoms that they study in their laboratories, and of no others, whether unknown or subservient to entirely different and incomprehensible laws. (Such differences of behaviour as have been observed are completely accounted for by the special conditions prevailing in the stars; these physics can evaluate and take into account.) This result is, moreover, of no less significance from the purely philosophical point of view.

It is not till we come to biology and medicine that we find the relevance of Astronomy beginning to wane. Admittedly, the study of at least one planet — Mars — has already provided strong grounds for presuming the existence of extra-terrestrial life; and this fact is obviously of the greatest interest when we come to consider the origin of life on the Earth. This, in turn, is largely dependent on the facts and theories of astronomy concerning the origin of planetary systems in general, and of the early history of our own Earth in particular. This fascinating problem has recently, in the United States provided astronomers and biologists with a common meeting ground — for the first time, possibly, but assuredly not for the last.

Of more immediate interest — and here we come back to practical preoccupations — are the effects that cosmic influences have on living organisms: not, of course, the imaginary influences to which astrologers try to hitch their pseudo-scientific vapourings, but the very real ones exercised by the Sun's radiation. These are complex and for the most part still imperfectly understood, although the diurnal and seasonal variations of this intense radiation — as well, no doubt, as the long-term variations associated with the Sun's own cycle of activity — do affect, in the most obvious manner, a wide range of vital processes: the growth of plants, the activity of certain bacilli and viruses (producing recurrent fevers), and general human metabolism.

Even leaving out of account many unexplained or unknown astronomical phenomena, it appears probable that doctors will come to concern themselves to an ever increasing extent with the data provided by astronomy. In the same way, meteorologists are beginning to take notice of the findings of astronomy, not only from the point of view of general statistics, but even in connection with their day-to-day forecasts.

To quote a single example in this connection, we might mention the interesting attempts that have recently been made in France to apply optical methods to the sounding of the Earth's atmosphere — methods based on the observation of the twinkling of the stars — to the forecasts of weather conditions for aircraft.

But these are fields which have not yet been opened up, and we are here concerned with presenting an account of the already magnificent achievements of astronomy.

We have yet to mention what — to really emancipated thinkers — constitutes astronomy's chief claim to glory, and its greatest service to mankind: its philosophical significance and its spiritual value.

Astronomy, which is without doubt the oldest of the sciences, has contributed more effectively than any other to the development of human thought. Although born of everyday needs (the calendar, the time, navigation) and the fears of primitive man in the face of the terrifying phenomena of Nature — for which reason it has remained closely associated with the superstitions of the astrologers in the minds of the uneducated, even, alas, to the present day — astronomy was nevertheless the first science to form the concept of natural physical law, which was demanded by the regularity of heavenly phenomena. It was also the first to develop, at a relatively early epoch, a discipline of accurate observation — to which the tablets of ancient Assyria and Chaldea, and the pyramids of Egypt, bear witness.

At a much later date, it was the Copernican revolution in astronomy that liberated men's minds from baseless superstition, and gave birth to the authentic scientific attitude and the rational outlook of the present day.

Thanks to astronomy, the enquiring spirit of man has at last discovered the true status of humanity in the Universe: a mere atom, but a thinking atom, situated on a microscopic planet, one of several revolving about a small and commonplace star, itself indistinguishable from a hundred thousand others, in the heart of a galaxy which in turn is lost among the millions that populate the tiny corner of space that we have been able to explore.

This atom, this man — you, in fact — far from feeling crushed and lost in the midst of this immensity, should on the contrary feel the less bewildered simply because he has been able to explore it, and even to begin to understand it. Realising this, he will be in a better position to appreciate the spell with which astronomy binds its devotees: the fascination and the wonder, no to be put into words, of the contemplation and the understanding of the heavens.

INDEX

(Page numbers in italic refer to tables)

A

Abaxial aberrations 456
Abbot 283
Absorption, atmosphere 98; interstellar 381–2
AC 70° 8247, *343*
AC +79° 3888, *310*
Achernar *310*
Achromatic lens 452
Adams 57, 198, 311, 494
Adonis 207
Adrastea 216
Aerolites, *see* Meteors
Air 85, 88, 91–2; electrical properties 103, 105; *see* Atmosphere
Airy 58
Albedo, definition 185
Aldebaran, colour 338; elements *310*; magnitude 305, 323, 325; naming 302; position 306; spectral type 342
Algol 318–9
Allegheny Obs. *455*
Aller 416
Altair, elements *310*; galactic longitude 360; magnitude 336; naming 302; spectral type 340
Altitude, affecting brightness *305*; record 90
Amalthea 216
American Association of Variable Star Observers 324
Ammonia, on Jupiter 212; Saturn 221; Uranus 226
Amphidromic systems, *see* Tides
Analytical Photography 477
Anaxagoras 148
Anderson C. 357
Andromeda, Great Nebula, *see* M 31
Andromedae, γ, *313*
Andromedae, δ, 319
Andromedae, ν, 400
Andromedae, R, 342
Andromedae, S, nova, 332
Andromedae, Z, 328
Angstrom unit 479
Annelid worms 174
Annual equation, Moon's 122
Antapex 308
Antares, colour 338; density 343; diameter 343; elements *310*; galactic longitude 360; mass 343; size 338; spectral type 342; super-giant 345
Antoniadi 186, 197–9, 204, 289
AOe 17415–6, *310*
Apennine Chain, Moon's 129
Aperture 453
Apex 308; galactic 393; Sun's classical 393
Aphelion 14
Apogee 14
Apollo 207
Apparition 192
Appulse 168
Apsides, *see* Progression
Aquarii, R, *322*, 328
Aquarii, CY, 320, *321*
Aquila, dark nebula *373*
Aquilae, α, elements 310; galactic longitude 360; spectral type 340; *see* Altair
Aquilae, η, *321*
Aquilae, R, *322*
Aquilae, Nova III, *326*
Arab astronomers 318
Arago 275
Arc, of longitude and latitude 80
Archimedes 283
Archimedes, crater 133
Arcturus, colour 338; elements *310*; magnitude 325; position 306; spectral type 342
Arend-Roland, comet 229
Argelander 318
Argelander's Star Catalogue, *see* Bonner Durchmusterung
Argon, in air 88
Ariel *225*
Arosa Obs. 475
Ashbrook-Jackson, comet *232*
Asideres 256
Asteroids, atmosphere 210; diameters, 63, 208; discovery 57, 205-6; distribution 59, 207; mass 210; nature 208–210; number 60, 206; orbits 59, 61, 207; origin 208–210; total volume 210
Aston F. W. 357
Astrographic Catalogue 304, 459, 465
Astronomical Unit (A. U.) 62, 462
Astrophysics 4, 177
Atlas, lunar 139
Atmosphere, of Mars 111; Mercury 111; planets 111; *see* planets, Stars, Sun
Atmosphere, Earth's 3, 6, 84, 283; absorbing light 98; affecting eclipses 159–160, analysis 160; basic component 82; colour 96; composition 85; density 85, 98; diffusing light 85, 96; electricity 104; fall-off 89; flattening 85; friction 87; height 85; 89; indispensable to life 84; ionisation 89; layers 86, 89, 99, 101; limit 86; mass 88; obstacle to extreme wavelengths 339; pressure 88; refracting light 98; structure 88; temperature 85, 90; thickness 85–6; tides 174; weight 88; *see* Air, Clouds, D layer, E layer, F layer, Ionosphere, Refraction, Stratosphere, Tropopause, Troposphere
Atmosphere, Moon's 111, 137–8, 148, 151
Atmospheric absorption 305
Atomic energy, *see* Stars, energy sources
Atomic mass, unit 357
Atoms, electron structure 482–4; in atmosphere 90; ionised 485, 496; special behaviour 492
Attraction, universal, *see* Gravity
Aurigae, α, diameter *338*; elements *310*, *315*; galactic longitude 360; spectral type 342; *see* Capella
Aurigae, β, *315*, *319*; A, *338*
Aurigae, ε, elements *313*, *319*; size of primary 337–8; A, *338*; B, *338*
Aurigae, ζ, comes 343; elements *319*; spectral type 343; super-giant 345; transparency 343; A, diameter *338*
Aurigae, T, nova, *326*
Aurigae, RT, *321*
Aurigae, UU, *325*, *326*
Aurorae 87, 89, 102–3; Australis 102; Borealis 102; on Venus 190; permanent general 296; Sun's part 286, 288; Zodiacal light 296
Axis, geographical 102; magnetic 102; of orbit 14

B

Baade, Walter 322, 329, 333, 407, 435, 444
Babcock H. W. 416, 494
Babinet 480
Bailey S. I. 383
Bailly, crater 132, 151
Baily 275
Baily's beads 166
Baize P. 348
Baldet F. 234, 237, 248
Balmer 482
Balmer's hydrogen lines 278, 482, 486
Barbier D. 160, 350, 379, 496
Barnard 329, 371
Barnard's Star, density 343; elements *310*; magnitude 306; mass 343; proper motion 306; radius 343; spectral type 342
Barometric pressure 88
Baum 145
Bayer 302
B. D., *see* Bonner Durchmusterung
BD + 5° 1668, *310*
BD +20° 2465, *310*
BD +43° 4305, *310*
BD −12° 4523, *310*
Becker 251
Becquerel, Henri 356
Beehive Cluster, in Cancer 380
Beer 145
Belopolsky 189
Belt of totality, *see* Eclipses
Berget, Prof. 83
Berlin Obs. 58
Besançon Obs. 498
Bessel 311, 316
Betelgeuse, elements *310*; period 325; size 338; spectral type 342; super-giant 345
Bethe, Hans 357, 359, 439, 445
Bethe's cycle, *see* Carbon-Nitrogen cycle
Bianchini 188
Bidault de l'Isle G. 289
Biela's Comet, disintegration 240, 250; elements *232*
Binaries, *see* Stars, Binary
Biot 258
Birefringent micrometer 463
Birkeland 103, 288
Bischoffsheim Obs. *454*
Bjerkness' sunspot theory 279–280
Black body radiator 494–5
Blink microscope 466
Bloemfontein Obs. *454*
Blue coal 174
Bode 62
Bode's law 58, 62, 436–7
Bohr, Niels 482–3
Bok B. J. 373, 430, 432
Bolides, *see* Meteors
Bolometer 470—471
Bolometric magnitudes 348
Bond G. P. 465
Bonner Durchmusterung (B. D.) 302–3
Boötes, cluster of galaxies *419*
Boötis, α, diameter *338*; elements *310*; spectral type 342
Boötis, η, 323
Bore, tidal 174
Borelly, comet *232*
Bosler γ, 105
Bosque Alegre Obs. *455*
Boss L. 392
Bowen I. S. 278, 369
Bowie W. 112
Bradley's Star Catalogue 302
Brahe, Tycho 318, 333; star catalogue 302; supernova 474
British Astronomical Association 145; Variable Star Section 324
Brooks II, comet *232*
Brorsen, comet *232*
Brorsen-Metcalf, comet *232*
Brückner, Prof. 289
Buffon 436
Buisson 467
Bunsen 481

C

Cabannes 296
Calcium, in Sun 273–8; on Earth 111
Calculated diameters 337
Callisto 215, *216*
Cambridge 405, 475
Cambridge Obs. *456*
Camelopardalis, Z, 330, *331*
Camus J. 188
Canals, Martian, *see* Mars
Cancer, cluster of galaxies *419*
Cancri, ζ, 314; A, B, *315*
Cancri, RS, *325*
Canis Majoris, α, elements *310*; galactic longitude 360; spectral type 340; *see* Sirius
Canis Majoris, β, 323
Canis Minoris, α, elements *310*; spectral type 340; *see* Procyon
Canopus, elements *310*; magnitude 304; spectral type 340
Canum Venaticorum, α, *313*
Capella, colour 338; elements *310*; galactic longitude 360; magnitude 304, 325; photosphere 350; separation 463; spectral type 342; surface gravity 350; temperature 350; A, elements *348*; A, spectrum 496
Capricorni, β, 315
Capricorni, δ, 58
Carbon, in atmosphere of variable stars 327; in comets 240; in solar atmosphere *351*; on Earth 112
Carbon dioxide, in air 88; in atmosphere of Earth 189, of Mars 198, of Venus 189, 190
Carbon monoxide, in meteorites 256
Carbon-Nitrogen cycle 357–358, 439
Carbon stars 342
Cardan 240
Carina, nebulous region 327
Carinae, α, elements *310*; spectral type 340; *see* Canopus
Carinae, η, nova, *326*, 330
Carinae, L, *321*
Carinae, R, *322*
Carinae, S, *322*
Carrington 286
Carte Photographique du Ciel 304, 311, 459
Casatus, crater 133
Cassegrain 454
Cassini G. D. 188, 190, 211, 275
Cassini J. J. 188
Cassini's Division 221
Cassiopeiae, α, 326
Cassiopeiae, γ, 328; spectral type 340, 497
Cassiopeiae, η, *313*, 315
Cassiopeiae, μ, 306
Cassiopeiae, π, 319
Cassiopeiae, σ, *325*
Castor, 314; *see* Geminorum, α
Cat's-eye photometer 463–64
C. D., *see Cordoba Durchmusterung*
CD − 37° 15492, *310*
CD − 44° 11099, *310*
CD − 46° 11540, *310*
CD − 49° 13515, *310*
Centauri, χ, elements *315*; first magnitude 336; motion 306; spectral type, 306; spectral type, mass, luminosity 343; A, elements *310*; A, spectral type 342; B, elements *310*; B, spectral type 342; C, elements *310*; C, see Proxima
Centauri, E, diameter *338*
Centauri, β, *310*
Centauri, ω, 382–3
Centauri, R, elements *322*, light curve 324
Centauri, T, *325*
Centauri, SX, 325
Centrifugal force 78; affecting gravity 82; Earth's shape 78; effect on Jupiter 210, 212
Cephei δ, 320, *321*
Cephei, μ, *325*
Cephei, T, *322*
Cephei, VV, 319, 320; elements *319*; energy output 440; size of primary 337–8; temperature 345; A, diameter *338*
Cepheids, brightness 346; categories 383; classical 321–2, 383, 440; cluster variables 321–2, 383; dwarf 323; elements *321*; long-period 321–4; magnitude 322; observations for extragalactic nebulae 411–412; radial velocities 393; reactions 359; spectra 490–1; spectral type 346; variables 320–4
Ceres 62–3, 205, 208
Ceti, o, diameter *338*; elements *322*; long-period variable 323; naming 318; spectral type 342; *see* Mira
Ceti, τ, *310*
Ceti, T, *325*
Chadwick J. 357
Challis 58
Chalonge D. 160, 351, 379, 496
Chandrasekhar 352–3, 355, 432–3, 443–5
Charcoal meteorites 256
Chladni 258
Chondritic structure, of meteorites 256
Chree 102
Chromatic aberration 452–4
Chromosphere, *see* Stars, Sun
Circular velocity 230

Clark, Alvan 316
Clavius, crater 132, 151
Clouds, on Earth, condensation 92–4; distribution 95; electric field 104; formation 92; heights 94; seen at distance 78; systems 95, 104; types 95; updraughts 104
Clouds, on Sun, *see* Sun, flocculi
Clouds, on Venus 187
Cluster, moving 308
Cluster variables, *see* Stars
Clusters, galactic or open, 380; density 380; 432; distances 380–2; equilibrium 434; Great Bear 380; Hyades 380; lifetime 432; main types 381; number of stars 380; others 380; Pleiades 380; proper motion 380; star types 381; visibility 382; *see* Nebulae
Clusters, globular, clue to Galaxy 385; composition 444; density 383–4; description 382; diameter 383; distance 383, 385; distribution 382, 385; examples 382; galactic concentration 382; luminosity 382; number 382; radial velocity 393; separate Population 444; shape 383; star types 383; variables 383
Clusters of galaxies, *see* Extragalactic Nebulae
Coal Sack Nebula 371
Cobalt, in meteorites 256
Coblentz W. W. 200, 470
Colour 96–7; *see* Spectrum, Stars
Colour equation 305
Colour excess 373, 377, 381–2
Colour index 305; of stars 339–340, 476; international 476
Coma, cluster of galaxies *419*
Coma, mirror aberration 455
Coma, of comets 234, 237
Comae, nova, *326*
Comae, 42, 314, *315*
Comas-Sola, comet *232*
Comes, definition 314
Comet of 1811, 236
Comet of 1886, 237
Comets, appearance 7, 234, behaviour near Sun 237, 240; captures 231; collision with 241–2; composition 237, 240; constitution 234; density 237; discovery 232; disintegration 240; disturbance 237; Earth passing through tail 242; elements of lost periodic *232* and periodic *232*; luminosity 240; mass 237, 241; names 234; number 232–3; observation 231, 233; old descriptions 241; orbits 19, 20, 229; periodic 230, 233; perturbation 231; relation to meteor swarms 250, 258; sighting 233; size 236; structure 236; tail 237, 240, 243; temperature 240; temporary 230; transparency 236; turbulence 240
Commonwealth Research Organisation, Sydney 392
Compass, magnetic 101
Comte, Auguste 478
Conjunction 12, 66; inferior, superior 66, 181
Constant of recession 322
Constellations, appearance 20; motion 308; nature 20, 301–2; number 302; use 302
Continents, of Earth 82–4
Convective equilibrium 353
Converging lens 452
Co-ordinates, celestial 302; Right Ascension and Declination 52; terrestrial 80
Copernicus 181, 385
Cordoba Durchmusterung (C. D.) 303
Cordoba Obs. 391
Core, Earth's 106; behaviour 110–11; metallic state 111; pressure 109; temperature 109
Cornu 489
Corona, *see* Stars, Sun
Corona, atmospheric 95
Corona Australis, nebulous region 327
Corona Borealis, cluster of galaxies *419*
Coronae, α, *319*
Coronae, R, 327, 346
Coronae, T, 326, 328, 330
Coronagraph 475–6
Coronal jets, *see* Sun
Coronium 278, 485
Corpuscular radiation 286, 288, 296
Cosmic cloud 297
Cosmic dust 375
Cosmic rays, energy 433; in Earth's atmosphere 89; origin 433
Cosmic static 390
Cosmic year 394
Cosmogony 427, 436
Cosmological constant 424, 434
Couder A. 491
Counterglow 294
Cowling 353
Crab Nebula 334, 474
Craters, lunar, appearance 129, 132–4, 139, 149; Archimedes 133; Bailly 132, 151; Casatus 133; Clavius 132, 151; comparison with volcanic craters 133, 139, 140; formation 129, 138; Grimaldi 149; height of walls 133; Messier 145; names *144*; nature of walls 142; Newton 133; number 133; Pickering 145; shapes 139; size 132, 149; Theophilus 133; theories 139; white rays 134
Crêpe Ring, *see* Saturn's rings
Crimean Astrophysical Obs. 190, *455*
Crommelin, comet *232*
Crucis, α, *310*
Crust, Earth's 105–6; analysis 111; earthquakes 108; varying density 112
Crux, dark nebula 370; galactic longitude 360
Cuffey 381
Cunningham 207
Curtiss H. D. 401
Cyanogen, in comets 240
Cygni, α, elements *310*; galactic apex 393; galactic longitude 360; spectral type 340
Cygni, β, *313*
Cygni, χ, *323*, 323–4
Cygni, P, spectrum 497; variability 329
Cygni, Q, nova, *326*
Cygni, R, 342
Cygni, W, *325*
Cygni Y, energy output 440; A, B, diameter *338*; A, B, elements *348*
Cygni, SS, elements *331*; variable type-star 330
Cygni, V 380, A, *338*
Cygni, 61, mass and luminosity 345; measurements 461; motion 306; parallax 311; planetary comes 317; spectral type 342, 345; system 317; A, elements *310*, *348*; B, elements *310*
Cygni, Nova III, *326*
Cygnus, rift in, 371

D

D layer, in atmosphere 90
Daniel, comet *232*
Danjon A. 160, 188
Danjon's micrometer 460
Danjon's photometer 463–64
Danvillier 296, 438; theory 287
Darkness 290–1
D'Arres, comet *232*
Davida 208
Dawes 204
Day, polar, length *46*
De Bort, Teisserenc 90
De Broglie, Louis 483
Declimation (Dec) 50, 302
Declination axis, on telescopes 457
Declination, magnetic 101
Degenerate matter, *see* Stars, white dwarf
Deimos 205
De Kérolyr 401
Delmotte G. 139
Delphini, γ, *313*
Dembowska 208
Demeter *216*
Deneb, elements *310*; galactic apex 393; galactic longitude 360; spectral type 340
Densitometer, *see* Microphotometer
Density, of Earth 81–2, 110, 137; of matter in space 422; of Moon 137; of planets *63*; of stars, *see* Stars
Desguin 438
De Sitter 424, 434
Deslandres 103, 273, 288, 474, 486
De Vaucouleurs 197
De Vico 188
Diameter, angular 19
Diameter, apparent 18
Diameter of planets *63*; of Stars *338*; of Sun, *see* Sun
Diffraction 454; spectra 479
Diffusion of light 85, 96–7
Dimensions of planets *63*
Dione *222*
Distance of planets *62*
Distances, astronomical 17, 298
Diurnal parallax 119–120
Diverging lens 451
Doerfels Mountains 129
Dollfuss, Audouin 146, 184, 186, 188–9, 204, 475
Dollond J. 452
Dominion Obs., Victoria *455*
Donville Obs. 256
Doppler 487
Doppler effect 307, 313, 426, 487–8, 492, 494
Doradus, β, *321*
Double stars, *see* Stars, binary
Draper H. 481
Draper W. 480
Dreyer J. L. E. 400
Dufay 292, 296
Duncan J. C. 401
Dunham 198
Dunlap Obs. *455*
Dust, affecting electric field 104; diffusing light 97
Dwarf stars, *see* Stars
Dwingeloo, radio-telescope 472

E

E layer, in atmosphere 90
Earth, a planet 6, 72; age 110, 427–430; antipodal point 73; atmosphere, *see* Atmosphere; brightness *464*; climate 284–5; composition 110–1; cooling 110; crust 83, 105; density *63*, 81–2, 110; diameter *63*, 72, 79; dimensions *79*; disturbance of ionosphere 288; effect of Sun 259, 281, 283 of solar cycle 288–9; elasticity 106; electrical field 101, 104; elements *62*; elliptical meridian *79*; equatorial radius *79*; evolution 437–8; flattening 78–9; future 446–7; importance 71; interior 106, 108, 111–2; land and sea 82, 84; length of equator *79*; loss of light elements 111; heights 83; history 146; magnetic storms and aurorae 286–8; magnetism 101–3; mass *63*, 81; metallic state 111; meteorological cycles 298; movement affecting observation 69, 71; oblateness *79*; orbital velocity 307, 393; original nebula 111; origin of life 438–9; place amongst planets 62, 71; place and movement 4; polar radius *79*; prehistory 438; radio-activity of crust 110; radius *79*; relative positions to Moon 121; resemblance to Mars 191; to Venus 186; revolution period 69, 71; rotation 64, 80, affecting eclipses 162–3 tides 172; rotational velocity in Galaxy 472; satellite the Moon 60, 113; sea beds 137; seen from Moon 151, 154; shadow 156, 158; shape 72, 76, 78–9; slowing by tides 174; structure 82; surface area *79*, 82; surface speed 80; temperature 106–7, 109, 111: tidal relation with Moon 118; vegetation, dependance on Sun 285 to 6; volume *63*, 79, 81
Earthlight 116, 124
Earthquakes 108–9
Ebro Obs. 175
Eccentricity of orbits 14; of binary stars *315*; of planets *62*
Eclipses 12, 121, 154–7
Eclipses, lunar, brightness 159–161; colour 159, 161; definition 156; duration 159; frequency 156–7; information from 160; mid-eclipse 159; penumbral; 158–9; total and partial; 156–9; visibility 157
Eclipses, solar, annular 162, 166; belt of totality 162–3; definition 156; duration 162, 166, 168; effect on animals 167; factors 161; frequency 156–7, 168; illumination 167; information from 167; mid-eclipse 166; pearl ring 166; penumbral belt 166; size 162–3, 166; total and partial 156–7, 161–4, 166; visibility 157, 162, 164
Eclipsing variable stars *319*
Ecliptic 31; relation to Moon's orbit 120–1 to planets' orbits 61; Zodiacal light 292
Eddington 323, 348, 352–4, 491
Edinburgh Obs. *456*
Edlen 278
Egypt, 74, reflector *455*
Einstein 356, 423–4, 434, 499; effect 494, 499
Electric field, of Earth 101, 104
Electrical measurements 467
Electro-magnetic disturbances, on Sun 276
Electro-magnetic waves, in atmosphere 89
Electrometer 469
Electron gas, see Fermi's gas
Electron multiplier 470
Electron pressure 496
Electronic telescope, *see* Telescopes
Electronics, applied to astronomy 470–1
Elements, abundance 432; formation 433; heavy 445
Ellipse 14, 229
Ellipsoid of revolution 78–9; relation to geoid 82
Ellipsoidal variables, *see* Stars
Elongation 66; of orbit 180
Elvey 375
Emden 280, 352
Emersion, *see* Occultations
Enceladus *222*
Encke's Comet *232*
Encke's Division 221
Energy, *see* Stars, Sun
Energy effect 426
Ephemerides 166
Equation of time, *see* Time
Equator, galactic 386
Equatorial mounting, *see* Telescopes
Equinox, vernal, coincidence with Moon's nodes 122
Equulei, δ, *315*
Eridani, α, *310*
Eridani, ε, 306, *310*
Eridani, o 2, motion 306; spectral type and luminosity 345; A, elements *310*; B, elements *310*; C, elements *310*
Eridani, 40, white dwarf 344; B, absolute magnitude *343*; B, diameter *338*; B, luminosity 348
Eros 207, 210; parallax 461–2
Eunomia 208
Europa *216*
Evaporation 92
Evection, Moon's 122
Evening Star, *see* Venus
Everest, Mount, height 83
Expanding universe, *see* Universe
Extrafocal photometry 466–7
Extragalactic distance 322
Extragalactic nebulae, absolute magnitudes 419; absorbing bands 407; apparent distribution 375; appearance 391, 401, 415; brightest stars 412, brightness 405–7; catalogues 400; characteristics 404; classification 401, 403; clouds of 421; cluster in Coma 419, in Virgo 419; clusters 421, 419, 421; colour 395, 398; colour-index 411; corrections 400; counts 400–1; definition 375; density 434; diameters 404–5; dimensions 412, 415; direction of rotation 417–8; discussion 395; distances 411; distribution 400; in clusters 432; of star types 407; dwarf galaxies 415, 419; elements of brightest *399*; evolution 446; extent 405; flattening 401, 415; formation 434–5; frequency of isolated and multiple 418; galaxies in Local Group *418*; globular and elliptical 401, 404–7, 411, 415, 419; groups 418–9; intergalactic bridges 421; irregular 403, 415, 419; luminosity 412; magnitude 412; mass 416, 434; mass of clusters 419; multiple systems *418*; nature 395, 399; nomenclature 399; principal clusters *419*; radial velocities 393, 416, 419, 422, 424; recession 422, 424; red shift in spectra 422, 424, 426; rotation 415–6; rotational velocities 416; separation 434; spectroscopic types 411; spiral and elliptical 446; spiral, normal and barred 403, 411, 415, 419; star types 407, 411; visibility 399–400; visible stars 407; *see* Local Group
Extragalactic supernovae, *see* Supernovae

F

F layer, in atmosphere 90
Fabricius 318
Fabry 263, 292, 468; photometer 464–7
Faculae, *see* Sun
Faye 278, 436
Faye, comet *232*
Fehrenbach C. 477, 491
Fermi's gas, electron gas 354; in contraction of stars 444; in Sun's corona 277
Ferrous particles, as meteoric fall-out 256
Figure of equilibrium 78
Filaments, *see* Sun
Filar micrometer 459
Filters, monochromatic 476–7
Finlay, comet *232*
Fireballs, *see* Meteors
Fixed stars 306
Fizeau 487
Flagstaff Obs. 198, 204
Flammarion, Camille 198, 289

Flares, *see* Sun
Flocculi, *see* Sun
Focal length 453
Focal plane photometry 465–6
Focal rations 453
Focus, of elliptical orbit 14
Fomalhaut *310*
Forbes, comet *232*
Forbidden lines, of spectrum 368, 483, 485, 497
Forbidden radiations 278, 471
Formaldehyde, in Venus's atmosphere 189
Fornax, nebular system *418*, 419
Foucault 456; pendulum 25–7
Fournier 197, 204
Fowler 354
Fraunhofer 479
Fraunhofer lines 278, 479, *480*, 489, 496
Front, meteorological 95

G

g, *see* Gravity
Galactic clusters, *see* Clusters
Galactic concentration *361*
Galactic equator 386
Galactic latitude 360
Galactic longitude 360
Galactic nebulae, *see* Nebulae
Galactic plane, *see* Galaxy
Galactic pole 360
Galaxies 332; evolution 446; Galactic pole 360
Galaxies 332; evolution 446; formation 434–5; see Extragalactic Nebulae
Galaxy, The, annual angular rotation 394; apex 393; boundaries 395, 399; centre 385–6; comparison with extragalactic nebulae 391; cosmic cloud 375, 377, 379; diameter 22, 386; dimensions 415; distribution 361–3, 385–386; elements *418*; equator 385; escape velocity 393; extension 385–6; galactic plane 363–4, 385–6; gravitational field 393; local cluster 391; mass 393–4; median plane 391–2; nature 360; nucleus 386; number of stars 364, 367; position in Local Group 419; radio sources 474; rotation 392–4; shape 21; spiral nebula 391, 411; star density near Sun 390; star streams 392; structure 385–6, 391–2 shown by neutral hydrogen 391–2; thickness 386, 391; velocity 394; *see* Clusters, Interstellar gas, Milky Way, Nebulae
Gale 107
Gale, comet *232*
Galilean satellites, *see* Jupiter
Galileo 213, 266, 302, 451, 498
Galle 58
Galvanometer 469
Gamow, George 359, 435, 439–440
Ganymede 215, *216*
Gaposchkin C. 327, 333
Gaposchkin S. 320, 327, 333–4
Gas laws, *see* Stars
Gegenschein 294
Gemini, cluster of galaxies *419*
Geminorum, α, *313*, 314; C, *315*; *see* Castor
Geminorum, β, *310*, 325; *see* Pollux
Geminorum, ζ, *321*
Geminorum, U, 330
Geminorum, Nova II, *326*
Geodesic Service of USA 112
Geoid 82
Geothermic degree 109
Giacobini-Zinner, comet *232*; passing 250
Giant stars, *see* Stars
Gill D. 462, 489
Gleissberg 276
Globular clusters, dimensions 415, 419; in Andromeda nebula 411; in Magellanic Clouds 411; like satellites 415; luminosity function 411; star types 411; stellar density 419
Globules, galactic, formation of stars 435, 440; magnitude 373
Goethe Link Obs. 381
Goldsmith 206
Goodricke 318, 320
Gould 391
Granite, in Earth's crust 111, 438
Gravity 14, 81, 112; acceleration of *g* 81, *82*; anomalies 112; centre of gravity 16; in universe 434; Newton's law 15; on Moon 136, 140, 154
Great Bear 306, 308, 380; *see* Ursa Major
Great Comet of 1843, 236; of 1861, 234, 236, 242
Great red spot, *see* Jupiter
Greenwich Obs. *454*, 498
Grigg-Mellish, comet *232*
Grigg-Skiellerup, comet *232*
Grimaldi, crater 149
Groombridge 34, A, *310*; B, *310*
Groombridge 1618, *310*
Groombridge 1830, 306
Gruithuisen 145, 149

H

Hades *26*
Hahn, Otto 357
Hale 273, 276, 278, 475 491, 493–4
Hall, Asaph 205
Halley 234, 241, 306
Halley's Comet 234, 242–3, 250; elements *232*
Hamburg Obs. *456*
Harding 205
Hargreaves's micrometer 460
Hartmann 375, 467; spectrocomparator 489
Harvard catalogues 400
Harvard College Obs. 251, 327, 330, 340, 367, 373, 383, 421, *455*, *456*, 463, 471; solar station 475
Harvard College Obs., Bloemfontein *455*, *456*
Haute-Provence Obs. *455*, 470, 491
Hayford 79
Head, of comet 234, 235
Heat, in atmosphere 92
Heating a planet 281
Heavenly bodies, types 6
Heaviside 89
Heaviside Layer 90
Hebe 208
Heidelberg Obs. 206
Heights, on Earth 83
Heisenberg's Uncertainty Principle 483
Heliostat 475
Helium, abundance 432; atomic weight 353, 357; discovery 357, 499; disintegration of uranium 428; in air 88–9; in formation of stars 445; of Sun 439; in Jupiter's atmosphere 212; in meteors 258; in solar atmosphere *351*; in Sun 111, 272–3, 277; in white dwarfs 355; liberation 428; preponderance in Sun and stars 351–3; spectrum 484; synthesis from hydrogen 357–8; particles 357–8
Helmert 79
Helmholtz H. von, 356, 359; mechanism 439
Hencke 206
Henrich 433
Henry, brothers 303
Henry Draper catalogue 340, 481
Hera *216*
Herculis, α, *313*, 325, 338
Herculis, δ, *315*
Herculis, κ, *325*
Herculis, g, *326*
Herculis, u, *319*
Herculis, DQ, nova, 317, *326*
Hermes 207
Herschel, John 390, 400
Herschel, William 57, 149, 187, 222, 302, 312, 361, 363, 382, 395, 400–1, 454, 478
Herschel-Rigollet, comet *232*
Herstmonceux (Greenwich) Obs. *455*
Hertzian waves 471
Hertzsprung *3*, *343*, 467
Hertzsprung-Russell diagram 343–6, 381, 432, 439–440
Herzberg 240
Hesperus, *see* Venus
Hestia *216*
Hidalgo 207
Hinks A. R. *462*
Hipparchus's Star Catalogue 302
Hodgson 286
Hoffmeister 251
Holmberg E. 418
Holmes A. 428, 430
Holmes, comet *232*
Horizon 73–6, *77*
Horizontal, variations 72
Horizontal depression 74, 76
Horse-head nebula 371
Hour angle 457
Hoyle, Fred 433, 435
Hubble, Edwin 368, 399, 401–5, 407, 411–2, 417–418, 422, 424, 426, 446
Hubble's Constant 322
Huggins W. 395, 481, 489
Humason, Milton 411, 422
Humboldt 395
Huygens, Christian 221
Huygens, Mount 129
Hyades, dissolution 432; motion 308; open cluster 380
Hyades 7, 16, *343*
Hydra, cluster of galaxies *419*
Hydrae, ε, A, B, *315*
Hydrae, R, *322*
Hydrae, U, elements *326*; spectral type 342
Hydrocarbons, in comets 240; in meteorites 256
Hydrogen, atomic weight 353, 357; conversion into helium 357–8; in air 88–9; in Earth's crust 112; in evolution of Sun and stars 439–440, in interstellar gas 375; in Jupiter's atmosphere 212; in meteorites 256; in solar atmosphere *351*; in Sun 111, 272–5, 277, 279–280; neutral, absorption by 351; neutral, distribution 391–2; neutral, radiation 471–2; preponderance in Sun, stars and interstellar matter 351–3, 428, 432; scarcity in white dwarfs 355; spectrum 482–4; stars 340
Hygeia 208
Hyperbola, orbit 14, 229
Hyperion *222*, 223

I

Iapetus *222*, 223–4
IC 10, *418*
IC 342, *418*
IC 1613, 412, *418*, 419
IC 4182, supernova *333*
Icarus 207
Illumination, of Moon 118, 122, 125, 128, 143, 145–6; of planet 281
Image converter 471
Immersion, *see* Occultation
Inclination, magnetic 101
Inclination, of planets *62*
Index Catalogues (I. C.) 400
Indi, ε, 306, *310*
Infra-red photographs, of Mars 199; of Venus 189
Infra-red radiation 478
Institut d'Astrophysique de Paris 334, 379, 421
Interferometer 313; separation of double stars 460–1
Intergalactic matter 421
International Astronomical Union 489–490
International Bureau of Weights and Measures, standard metre 79
International polar sequence 490
International Time Bureau 498
Interstellar absorption 381–2
Interstellar cloud 375, 379, 381, 386, 394
Interstellar dust 375, 379
Interstellar gas, absorption 375–7; composition 375; cosmic state 390; hydrogen 375; mass 379; rotation 394; spectroscopic evidence 375
Interstellar matter 407
Io 214, *216*
Ionisation, atmosphere 89, 105; definition 484; ionosphere 288; lightning 104; meteor trains 248; Sun's corona 278; temperature 496
Ionised calcium stars 340
Ionised carbon monoxide, in comets' tail 240
Ionised helium stars 340
Ionised nitrogen, in comets' tail 240
Ionosphere 89; ionised layers 90; radar 248; solar eclipses 167; Sun's effect 288
Ions, atmosphere 89; electric field 104
Iris 208
Iron, Earth's crust and interior 111; meteorites 256; Sun's atmosphere *351* and corona 278
Iron oxide, on Mars 200
Irradiation, of Moon 138
Island universes 395, 399
Isogonics 102
Isophotes 390
Isostasy 112
Isotherms 284
Isotopes 356; age of rocks 428; proportions 432

J

Jansky K. G. 471, 474
Jansen 466
Jeans, Sir James 404, 436–8, 446
Jeffreys H. 436
Jodrell Bank, radio-relescope 472, 473
Johannesburg Obs. *454*
Joliot F. 356, 434
Joliot-Curie I. 357
Joy 393
Juno 63, 205, 208
Jupiter, apparent diameter 18, 211; appearance 211; atmosphere 212, 215; belts 211; colour 211; composition 212–3; core of ice 213; density *63*, 210; diameter *63*, 210; distance from Earth 210 from Sun 210; eccentricity of orbit 210; elements *62*; equatorial zone 211; flattening 79, 210; Great Red Spot 211–12; illumination 213; inclination of axis 210; known of old 57; magnitude 304; markings 211; mass *63*, 210, 343; observation 178, 214; occultation 168; opposition 210–11, 214; orbit and motion 210; phase-effect 67, 211; physical condition 212; polar caps 211; pressure at centre 355; revolution period 210; rotation period 210, 212; satellites 60, 212–5; South Tropical disturbance 212; surface 212; surface gravity 210; synodic period 211; temperature 213; turbulence 211–12; volume 210
Jupiter's satellites, atmosphere 215; brightness 215; eclipses 214; elements *216*; occultations 214; orbits 213–5; rotation 215; size 215; transits 214–5

K

Kant, Emmanuel 361, 395, 436
Kapteyn 304, 392
Kapteyn's Star 306, *310*
Kapteyn's star streams 392–3
Keeler J. E. 401
Kelvin, Lord 106, 356
Kempf 463
Kenelly 89
Kepler, Johannes 16, 240, 275, 334
Kepler's Laws 16, 114, 314, 462
Kepler's Somnium 148
Kepler's Supernova 474
Kiloparsec, *see* Parsec
Kimball 263
Kinetic theory of gases, Earth's composition 111
Kirchhoff's Radiation Laws 262, 481, 483, 494
Kitts Peak Obs. *455*
Kivu Obs. *456*
Kopff, comet *232*
Koppen 289
Kostinsky effect 459
Kothari 355
Kozirev N. 190
Krakatoa, eruption 90, 107, 161, 186
Kraus J. D. 189
Kron 469
Krüger 60, luminosity and mass 345; spectral type 342, 345; A, diameter *338*; A, elements *310*, *348*; B, elements *310*, 343, *348*
Krypton, in air 88
Kuiper G. P. 189, 198, 215, 224, 227–8
Kulik L. 255

L

L 726–8, 317
Lacaille 8760, *310*
Lacaille 9352, 305, *310*
Lacaille's Star Catalogue 302
Lacertae, CP, nova, *326*
Lacertae, Nova I, *326*
Lacertae, Nova II, *326*
Lacroute P. 433
Laetitia 208
Laffineur M. 262
Lalande 59; Star Catalogue 302
Lalande 21185, *310*
Lallemand A. 469
Lamond 59
Land and sea 82–4, 91
Land pole 83
Landau 444
Lane H. 352
Langevin, Paul 356–7
Langley 470
Laplace 241, 436
Lassell 227
Latitude 80; lengths 80, *81*
Latitude, celestial, Moon 122; Sun 122
Latitude, galactic 360
Lavoisier 87, 256
Lead, from radio-active disintegration *428*
Leavitt, Miss 321, 383, 411, 444
Leibnitz Mountains 129
Lemaitre, Abbé G. 423–4, 426, 433–4, 445–6
Lenouvel F. 470
Leo, cluster of galaxies *419*
Leonids 250
Leonis, α, *310*, 476
Leonis, γ, *313*
Leonis, R, *322*
Lescarbault Dr 60
Le Verrier 57, 60, 62
Lévy, Michel 247
Librae, δ, *319*
Libration, Earth 151; effect of Moon's 138, 143; Mercury 183; Moon 119
Lick Obs. 381, 400–1, 416, *454*, *455*, 469
Life, on Mars 200; on Venus 189, 191
Light, speed 17, 298; *see* Absorption, Colour, Diffusion, Refraction
Light curve 319
Lightyear 18, 309
Lightning 104–5
Lindblad B. 393–4, 418, 446, 472
Littrow 149
Liverpool Tidal Institute 172

Local cluster 367–8, 391
Local Group, centre 419; definition 418; members *418*, 419; relation to Virgo cluster 419; volume 419
Lockyer N. 357
Loewy 139, 140
Lomonosov M. V. 189
Long time-scale, see Universe, age
Long-wave radiation, from Sun 286
Longitude 80; lengths *81*; see Meridians
Longitude, galactic 360
Lost period comets, see Comets
Lovell, Dr A. C. B. 472
Lowell Obs. 59, 418, *456*, 470
Lowell, Percival 59, 188–9, 204–5
Luminosity classes, of stars 311, 344–7
Luminosity function 336
Luminous layers, in Earth's atmosphere 296
Lunar month, *see* Lunation
Lunation 115, 122; Saros cycle 157
Lundmark 334
Luyten 745–6, *343*
Luyten 789–6, *310*
Lyman 482; hydrogen lines 482
Lynx, dark nebula 371
Lyot, Bernard 146, 189, 200, 204, 273–6, 475
Lyrae, α, elements *310*; spectral type 340; see Vega
Lyrae, β, 318–9, 325; elements *319*
Lyrae, R, *326*
Lyrae, RR, *321*, 322
Lyttelton 436

M

M 3 in Canes Venatici 383
M 11 in Scutum 380
M 13 in Hercules, 382–3
M 31, Great Nebula in Andromeda (NGC 224), axis 421; brightest stars 412; colour 398; composition 399, 407; description 400; dimensions 415; distance 322, 332, 399, 411; elements *399*, 418; Local Group 419; luminosity function 411; magnitude 419; nature of spirals 403; radio source 474; rotational velocity 416; variables 411; visibility 400
M 32 in Andromeda (NGC 221) description 400–1; elements *399*, *418*; Local Group 419
M 33 in Triangulum (NGC 598), angular displacement 417; brightest stars 412; density 416; description 400; distance 411; elements *399*, *418*; Local Group 419; mass 416; nature of spirals 403; rotation period 417; rotational velocity 416; variables 411
M 37 in Auriga 380
M 51 in Canes Venatici 477
M 56 in Lyra 383
M 59 (NGC 4621), 401
M 64 in Canes Venatici (NGC 4736), *399*
M 66 in Leo (NGC 3627), *399*
M 81 in Ursa Major (NGC 3031), elements *399*; nature of spirals 403
M 82 in Ursa Major (NGC 3034), *399*, 403
M 83 in Hydra (NGC 5236), *399*
M 101 in Ursa Major (NGC 5457), elements *399*; nature of spirals 403
McCormick Obs. *454*
McDonald Obs. 375, *455*
McLaughlin D. B. 330, 493
McMath-Hulbert Obs. 475
Mädler 145
Magellanic Clouds, brightest stars 412; Cepheid variables 321; distances 411; formation 446; satellites of Galaxy 419; type 403, 419; visibility 400; see Nubecula Major and Minor
Magma 106; of Moon 139–140
Magnesium, in Earth's crust 111; in meteorites 256; in solar atmosphere *351*
Magnesium chloride, in Venus's atmosphere 189
Magnetic axis 102
Magnetic compass 101
Magnetic declination 101
Magnetic field, of Earth 101–3, 105; spectroscopic perturbation 491–2
Magnetic meridian 101
Magnetic pole 102
Magnetic storms 101–2, 104–5; Sun's part 286–8
Magnetic variation 101
Magnetism, terrestrial 101
Magnification 453
Magnitude 476–7; absolute 348; bolometric 348
Magnitude function 364, 367; empirical 304; in Galaxy367
Magnitudes, stellar 304; absolute 311; catalogues 305; photographic 304–5; photovisual 305; visual 304–5
Main Sequence, definition 344; *see* Stars
Malus's Law 462
Maria, lunar, Mare Imbrium 132; names *144*; nature 124, 129, 140–1; origin 129, 136, 139–140; surfaces 137
Marius, Simon 400
Mars, angular diameter 462; apparent diameter 192; atmosphere 195, 197; bands of moisture 197; canals 195, 200, 204–5; charts 194; climate 195; clouds 197; colour 193, 197, 199; conjunctions 192; density *63*, 191; deserts 198, 200; diameter *63*, 191; distance from Earth 187, 192 from Sun 192; eccentricity of orbit 191–2; elements *62* 191; flattening 191; haze 197; inclination of axis 191–3; known of old 57; life 200; magnitude 304; markings 193, 195, 198; Martians 204; mass *63*, 191; mountains 199; naming surface features 194; oases 204; occultation 168; oppositions 71, 192, 200; orbit 192; orbital velocity 393; parallax 463; phase-effect 193; polar caps 194–196, 198, 289; relative to Asteroids 59; revolution period 191–2; rotation period 191; satellites 60, 205; seasons 191–2, 194, 196, 199; surface 197–200; appearance 193, area 191; surface gravity 191; surface pressure 197; temperature 197–8; vegetation 199; velocity 192; visibility 191, 193; volume *63*, 191; water 197–9; water vapour 198
Mass 14; of planets *63*; *see* planets, Stars, Sun
Mass-energy equation 356
Mass-luminosity relation *348*, 353–4, 440
Mayall 416
Mayer R. 190, 356
Measured diameters 337
Megaparsec, *see* Parsec
Menzel D. H. 189–191, 351
Mercury, albedo 185; angular separation from Sun 181; apparent movement 60, 66, 69, 180; atmosphere 185–6; aureole 185; colour 184; comparison with Moon 184–5; conjunction 181–2; density *63*; diameter *63*, 183; displacement of perihelion 499; distance from Sun 179; eccentricity of orbit 180; elements *62*; elongation of orbit 180–1; inclination of axis 183–4 of orbit 182; inferior planet 66; known of old 57; libration 183; mass *63*, 186; mountains 184; observation 57, 60, 66–7; orbit 60, 180–2; phases 66, 181, 183; reflection of light 185; rotation period 183; surface 184 186; temperature 185; terminator 184, 186; transits 182–3, 185; twilight 186; veils 186; velocity 180; visibility 181, 184; volcanoes 184; volume *63*; water 185
Meridian 80; geographical 101; lengths 80, *81*; magnetic 101
Meridian circle 458
Meridian photometer 463
Messier, Charles 334, 400; catalogue of nebulae and clusters 380, 400
Messier, crater 145
Metagalaxy 400, 421
Metallic state, of Earth's interior 111
Meteorology 84; phenomena in troposphere 90
Meteors and Meteorites, Adrar 255; age 258, 430; appearance 246–7; Arizona 252; brightness 246–7; colour 247; composition 111, 256; craters 139, 252, 255; definition 8, 245; designation 250; description 86; detection by radar 248; detonations 247; fall 251; fall-out 256; frequency 249; height of appearance 246; incandescence 246–7; ionised trails 472; Kenya 255; large 252; limit 90; mass 249, 252, 256; nebulous 248; number 249; orbits 249–251 velocities 251; origin 250–251, 256, 258; particles 255; periodic 250; perturbations 250; radiants 249–250; relation to comets 250, 258; showers 249–250, 255; Siberian 252, 255; temperature 247; train 246, 248; velocity 246–7, 251; visibility 245, 249, 251
Methane, in atmosphere of Jupiter 12, Neptune 227, Saturn 221, Uranus 226
Metre, definition 79
Meudon Obs, 200, 204, 476
Michelson 107, 337, 461
Michelson-Morley experiment 499
Michigan University Obs., *see* Bloemfontein
Micrometers 459–60
Micron 479
Microphotometer 467
Mid-eclipse, *see* Eclipses
Milan Obs. 200
Mile, *see* Nautical Mile
Milky Way, age 432; appearance 360; 385; brightness 360–1; distribution of stars 361, 363, 371; nature 10, 20–1, 361; number of stars 364, 367; position in sky 360; width 361, 386; *see* Galaxy, Nebulae
Miller 439
Mimas *222*
Mindanao Deep 84
Mineur H. 393, 432
Minor Planets, *see* Asteroids — 7° 3632 A, *343*
Mira, magnitude 322–3; period 322; size 338; spectral type 342; spectrum 497; supergiant 345; variable 318, 328; *see* Ceti, *o*
Mirages 100
Miranda *225*, 226
Mirrors 454–5
Mizar 313; separation 461
Molecules, behaviour in atmosphere 88–9; thermal agitation 92, 111
Molecular radiation, see Radiation
Monoceros, nebulous region 327
Monocerotis, U, *325*
Monocerotis, Nova I, *326*
Monocerotis, Nova II, *326*
Monochromatic polarising filter 476
Montaigne 190
Montanari 318
Moon, angular diameter 19, 113; annual equation 122; apparent diameter 18; appearance 126; atlas 139; atmosphere, *see* Atmosphere behaviour relative to Sun 124; brightness 116, 124; celestial latitude 122; clefts and fissures 137, 139–140, 151; colours 141–3; comparison with Mercury 185; composition 141; craters, *see* Craters; dead planet 148; deformation 119; density 113, 137; diameter 113–4; displacement 114, 120–2; distance 113–4; Earth's tidal influence 174; eccentricity of orbit 122; evection 122; gravity 136; 140, 154; height in sky 124; height of formations 124, 128; history 137, 139, 146; illumination 116, 124–5, 128, 143, 145; inclination of axis 119, of orbit 120–1; influence on Earth's atmosphere 174 and rainfall 175; internal heat 137; irregularities in behaviour 122; libration 119, 138, 143; life 138, 148; lunation 115; magma 139–140; maps 129, 139, 145; mass 113, 137; mineral composition 136–7; motion 7, 12, 114, 122 seen by solar eclipse 166; mountains, *see* Mountains; nature 7; nodes of orbit 156–8; observation 178; occultations 137, 168; old moon in young moon's arms 116; orbit 114, 122, 156–8, 162, 168; orbital velocity 114, 118–9; origin 148; parallax 309; phases 66, 115–6, 122, 124; physical condition 126; poles 119–121; principle formations *144*; progession of lines of apsides 122; radar contact 472; reflection of light 125; 185; revolution period 114; rotation 118; rotation of lunar sky 154; seas, see Maria; seen at Earth's poles 124; shadow 161–2; size 113–4; soundless 154; spectrum 160; Sun's effect during eclipses 289; surface 141, 146, 166; surface area 113; surface features 128, 137, 139, 149; surface temperature 146, 148, 154; synodic period 115, 170; tidal effect 118, 170; transit time 115; traveller 150; vegetation 149; velocity 12; visible area 118, 120; visibility 122, 124; volcanoes 140, 146; volume 113; water 128–9, 136, 146; white rays 134, 137, 140; *see* Earthlight, Eclipses
Moonlight 124–6
Morehouse, comet 240
Moreux T. 289
Morgan W. W. 347; diagram 346–7
Morning Star, *see* Venus
Motions, celestial, measurement 458
Mouchez, Admiral 303
Mouchot's solar boiler 284
Moulton 436–7, 446
Mount Palomar Obs. 322, 333, 419, 421, 426, 444, *455*, *456*, 477
Mount Stromlo Obs. *455*
Mount Wilson Obs. 59, 185, 189, 198, 329, 333, 375, 377, 383, 386, 394, 399 to 401, 405, 407, 411–2, 417, 419, 422, *455*, 461, 470, 477, 491, 494; solar tower 475, 493
Mountains, Earth's 77, 83
Mountains, lunar, Apennines 129, 132; appearance 128, 132, 136, 149; composition 139, 141; Doerfels 129; heights 128–9; Huygens 129; Leibnitz 129; massifs 132; names *144*; origin 140; relation to casters 139; Rook 129
Moving cluster 308
Moving clusters 380
Mrkos, comet 229
Muller 251, 463; micrometer 460
Multiple stars, *see* Stars
Muraour 247

N

Nautical mile 80
Nebulae; bright galactic 368; Coal Sack 371; diffuse 370; globules 373; Horse-head 371; nature 10, 297; spectrum 368; spiral 22, 331; velocity 12; *see* Interstellar gas
Nebulae, Dark Galactic 370–1; absorption 373, 377; composition 373; condensation into super-giants 440; density 373; dissipation 432; distance 373; mass 373
Nebulae, Extragalactic, *see* Extragalactic Nebulae
Nebulae, Planetary 369; density 370; distance 370; distribution 370; temperature 369
Nebular variables, *see* Stars
Nebulium 278, 369, 485, 497
Neon, in air 88
Neptune, apparent diameter 227; appearance 227; atmosphere 227; belts 227; colour 227; constitution 227; density *63*, 226; diameter *63*, 226; discovery 57, 59, 62; distance 226; eccentricity of orbit 226; elements *62*; flattening 226; inclination of axis 226 of orbit 226; magnitude 304; mass *63*, 226; opposition 71, 226; revolution period 226; rotation period 226; satellites 60, 227; specific gravity 226; temperature 227; velocity 226; visibility 227; volume *63*
Neptune's Satellites *227*
Nereid *227*
Netherlands Foundation for Radio-Astronomy 392
Neuchâtel Obs. 498
Neujmin I, comet *232*
Neujmin II, comet *232*
Neujmin III, comet *232*
Neutral helium stars 340
New General Catalogue (NGC) 400
Newcomb 292
Newton, Sir Isaac, 14, 96, 170, 241, 455, 478, 498
Newton, crater 133
Newton's Law 14–16, 312
NGC 55 in Sculptor, *399*
NGC 147 in Andromeda, *418*, 419
NGC 185 in Andromeda, *418*, 419
NGC 205 in Andromeda, *418*, 419
NGC 221, *see* M 32
NGC 224, *see* M 31; supernova *333*
NGC 253 in Sculptor, *399*
NGC 598, *see* M 33
NGC 891, 407
NGC 1003, suoernova *333*
NGC 2859, 403
NGC 3031, *see* M 81
NGC 3034, *see* M 82
NGC 3115 in Sextans, *399*, 401, 416
NGC 3379, 401
NGC 3627, *see* M 66
NGC 4039, 403
NGC 4273, supernova *333*
NGC 4303, supernova *333*
NGC 4321, supernova *333*
NGC 4424, supernova *333*
NGC 4449, 403; *see* Magellanic Clouds
NGC 4486, supernova *333*
NGC 4527, supernova *333*
NGC 4565, 407
NGC 4594 in Corvus, *399* 403, 416
NGC 4621, *see* M 59; supernova *333*

NGC 4631 in Canes Venatici, *399*
NGC 4725, supernova *333*
NGC 4736, *see* M 64
NGC 4826 in Coma, *399*
NGC 4945 in Centaurus, *399*
NGC 5128 in Centaurus, *399*, 403
NGC 5236, *see* M 83
NGC 5253, supernova *333*
NGC 5457, *see* M 101
NGC 5494, 407
NGC 5850, 403
NGC 6822 in Sagittarius, 403, 411–2, *418*, 419; *see* Magellanic Clouds
NGC 6946, possibly Local Group *418*; supernova *333*
NGC 7479, 403
NGC 7793 in Sculptor, *399*
Nicetus 241
Nicholson 189
Nickel, in Earth's interior 111; in meteorites 256; in Sun's corona 278
Night, polar 48; length *46*
Night light 290–2, 296–7; *see* Zodiacal light
Nitrogen, decomposed 90; in air 88–9; in meteorites 256; in solar atmosphere *351*; ionised, on Venus 190; size of molecule 94
Nodes, ascending, 61; coincidence of Moon's 122; definition 120; descending 61: of Moon's orbit 156–8; precession of Moon's 121; *see* orbits of planets
Northern Lights, *see* Aurorae
Nova Aquilae 1918, 329 – 330, 370
Nova Aquilae 1936, 330
Nova Aurigae 1892, 329–330
Nova Cygni 1920, 329 – 330
Nova Cygni 1942, 330
Nova Geminorum 1912, 330
Nova Herculis 1934, 330
Nova Lacertae 1936, 329 – 330
Nova Persei 1901, 329 – 330, 370
Nova Pictoris 1925, 330
Nova Puppis 1942, 330
Nova Sagittarii 1936, 330
Nova T Sagittae 1913 and 1946, 330
Novae, appearance 10; causes 331, 444–5; description 329; elements of brightest *326*; expulsion of gases 329; formation 444–5; in Andromeda 332, 411; magnitude 330; nebular shell 329; periods and amplitudes 330, *331*; rapid 330; recurrent 330; slow 330; spectra 491, 497; spectral type 346
Nubecula Major in Norma, *399*, *418*
Nubecula Minor in Norma, *399*,*418*
Nubecula Major and Minor, *see* Magellanic Clouds
Nuclear reactions, in stars 428
Nucleus, of comets 234–7
Number effect 424–6

O

Oases, *see* Mars
Oberon *225*
Objective lens 451
Observatories 458
Occultation, appearance 169; by Moon 137; definition 12, 168; emersion 168; immersion 168
Oceans, of Earth, 82–4
Ocular, lens 451, 454
Olbers 205
Olbers, comet *232*
Oort, Prof. 393–4, 407, 411, 416, 471
Oort's Constant 394
Open clusters, *see* Clusters, galactic
Ophiuchi, α *326*
Ophiuchi, RS, 330
Ophiuchi, RS, Nova III, *326*
Ophiuchi, 70, elements *315*; measurements 459; planetary comes 317; system 317; A, diameter *338*; A, elements *310*; B, diameter *338*; B, elements *310*
Ophiuchi, Nova II, *326*
Ophiuchus, constellation 302; dark nebula, elements *373*; supernova 334
Öpik 251
Opposition 66, 71
Optical doubles, *see* Stars, binary
Orbits 12, 14; of binaries *315*; of planets 61, *62*, 67; *see* individual bodies
Orion, moving cluster 308; nebula 327, 368, 395
Orionis, α, diameter *338*; elements *310*, *325*; spectral type 342
Orionis, β, elements *310*, *313*; spectral type 340
Orionis, γ, 340
Orionis, δ, elements *319*; spectral type 340; spectrum 375
Orionis, ζ, spectral type 340; nebula near 371
Orionis, χ, 379
Orionis, U, *322*
Orionis, W, *322*, 325
Oterma, comet *232*
Oxygen, atomic weight 357; decomposed 90; doubly ionised 369; in air 88–9; in Earth's atmosphere 160; in Earth's crust 111; in Mars's atmosphere 198; in meteorites 256; in solar atmosphere *351*; in Sun 272; in Venus's atmosphere 189–190
Ozone, in Earth's atmosphere 90, 160, 288

P

P waves, through Earth 110
Packing fraction 357
Pallas 63, 205, 208
Pan *216*
Paneth 430
Parabola, orbit 14, 229
Parabolic escape velocity 111
Parahelia 95
Parallax, definition 309; diurnal 119–120; dynamic 311; hypothetical 311; spectroscopic 311; trigonometrical 311
Paré, Ambroise 241
Parenago P. 331, 344
Paris Obs. 160, 303, *454*, 470, 498; *see* Meudon
Parsec 309
Parsons, William, see Rosse
Paschen 482; hydrogen lines 482
Patroclus 208
Pauli's Exclusion Principle 355
Pavonis, α 321
Pavonis, SX, *326*
Pearl ring, *see* Eclipses, solar
Peary 286
Pease F. G. 463
Pegasi, β, diameter *338*; rotation 494
Pegasus, cluster of galaxies *419*
Penumbra 158–9, 166; of sunspots 266
Perigee 14
Period-Luminosity relationship 321, 383, 411, 444
Periodic comets, *see* Comets
Perihelion 14
Perkins Obs. *456*
Perrin, Jean 357
Perrine, comet *232*
Persei, β, 318, *319*; A, *338*; see Algol
Persei, ζ, 379
Persei, η, *313*
Persei, ϱ, *326*
Persei, double cluster h and x 380
Persei, Nova II, *326*
Perseids 250
Perseus, cluster of galaxies *419*
Personal equation 305
Perturbation 17
Peters 316
Pettit E. 273, 470
Phases, see individual bodies
Phobos 205
Phoebe *222*, 223
Photo-electric, effect 470; measurements 469
Photographic measurement 461, 470
Photography, by moonlight 125; of Asteroids 206; of Stars 305
Photometers, astronomical 462–5
Photometric Catalogue 463
Photometric grid 467
Photometry 477
Photon, emission 483
Photosphere, see Stars, Sun
Photovisual magnitudes 305
Piazzi 62, 205
Pic du Midi Obs. 184, 186, 204, 453, 475
Piccard, Prof. 76, 90
Pickering E. C. 313, 463, 481
Pickering W. H. 149
Pickering, crater 145
Pictoris, RR, nova, *326*
Pisces, meteor in, 472
Piscis Australis, α, *310*; *see* Fomalhaut
Piscium, 19, spectral type 342
Planck, Max 482, 495
Planck's Constant *h*, 483
Planets, angular diameter 19; apparent brightness 304; apparent dimensions 64; apparent movements 67–68; appearance 7, 64; density *63*; diameter *63*; dimensions 62, *63*; direction of orbits 61; discussion 6, 57, 177; distance 62; 64; distribution 61–2; eccentricity 62; effect of refraction 99 of solar radiation 281; elements *62*; formations 436–8; giants and dwarfs 63; inclination 62; inferior and superior 66; known of old 57; mass *63*; measuring temperature 470; movement 7; number 59; observation 64, 67, 178; orbits 19, 59, 62; orbital planes 61; phases 64, 66–7; revolution period 62; symbols 62; unknown 60; velocity 12; volume *63*; *see* individual planets
Planetary comites 317
Planetary nebulae, *see* Nebulae, planetary
Planetary systems, to stars 317–8
Plaskett J. S. 393
Pleiades, density 380; disintegration 432; distance 380; life-time 432; motion 308; nebula 368; nebulosity 398; number of stars 380; open cluster 380; variability of stars 327
Plutarch 148
Pluto, albedo 228; atmosphere 228; brightness 228; density 228; diameter 228; dimensions 63; discovery 58–9, 62, 466; distance 227; eccentricity of orbit 227; elements *62*; inclination of orbit 227; magnitude 304; mass 228; orbit 20, 60; revolition period 228; volume 228
Poincaré, Henri 356
Polar axis, on telescope mountings 457
Polar day and night *46*
Polar lights, *see* Aurorae
Polar sequence 305, 490
Polaris 340
Polarisation of light, on Mars 198; on Venus 189
Pole, galactic north 360; land 83; magnetic 102; wandering of geographical 107 of magnetic 102
Pollux, elements *310*; magnitude 325
Pons-Brooks, comet *232*
Pons-Winnecke, comet *232*
Populations, stellar; *see* Stars
Pores, see Sun
Poseidon *216*
Positions, celestial, measurement 460
Potassium, on Earth 112
Potsdam Obs. *454*, 465
Praesepe Cluster, in Cancer 380
Precession, of Moon's nodes 121
Pressure, in Earth's interior 106, 109, 111
Primary colours 478
Prime focus 454
Prismatic spectrum 479, 491
Proctor 194
Procyon, companion of, 316; elements *310*; first magnitude 304, 336; spectral type 340; A, diameter *338*; A, elements *310*; B, absolute magnitude *343*; B, elements *310*; B, luminosity 348
Progression of line of aspides 122
Prominences, *see* Sun
Proper motions, *see* Stars
Proton 357
Prout Dr 356–7
Proxima Centauri 335–6
Psyche 208
Ptolemy 304; 'Almagest' 302
Puiseux P. 133, 139, 140
Pulkova Obs. *454*
Puppis, ζ, 340
Puppis, L, *322*
Puppis, V, *315*, *319*
Puppis, Nova, *326*
Purkinje effect 125, 291
Pyxidis, T, 330

Q

Quadrant, terrestrial 79
Quadrature, definition 66; tidal effect 172
Quantum Theory 424, 482, 495
Quartz monochromator 476

R

Rabe E. 463–4
Radar, contact with Moon 472; detecting meteor trains 248
Radcliffe Obs., Pretoria *455*
Radial velocity, definition 307; measurement 488 – 491
Radiant, of meteor swarm 249–250
Radiation, in atmosphere 91, 296–7; laws 482–5, 494; *see* Stars, Sun
Radiation pressure, on comets tails 240; on Sun 277
Radiative equilibrium 352
Radio-activity, discovery *356*; disintegration of elements 428, 433; Earth's crust 89, 105; Earth's internal heat 110; volcanic eruption 107, 110; *see* Stars, Sun, energy
Radio astronomy 471
Radio observation 262
Radio sources in Galaxy 474
Radio Sun 262
Radio-telescope, *see* Telescopes
Radio waves, affected by ionosphere 89
Radium, on Earth 110
Radium clock 356
Radius vector 17
Rainfall, causes 95; Moon's influence 175
Ramsay 88, 111
Rayleigh, Lord 88, 97
Rays, lunar, between craters 134; origin 140; relation to Maria 137
Reber G. 471
Recession of extragalactic nebulae 422
Recurrent novae, *see* Novae
Red dwarfs and giants, *see* Stars
Red shifts, Doppler effect 426; of nebular spectra 422, 424, 426
Redman R. O. 405
Reflectors, *see* Telescopes
Refraction of light, affecting visibility 99; at sunrise and sunset 100; at various heights *98*; by Earth's atmosphere 159, 160; calculation 99; effect on horizon 75–7; making errors 98
Refractive index 452
Refractors, *see* Telescopes
Regulus *310*; see α Leonis
Reinmuth 207
Reinmuth I, comet *232*
Reinmuth II, comet *232*
Relativity, theory 422–4, 494, 499
Resolving power, for different apertures *453*
Retrograde motion 70
Reversing Layer, *see* Stars, Sun
Revolution period 17; of planets *62*
Rhea *222*, 223
Riccioli 190, 313
Rice-grain, *see* Sun
Richardson R. S. 189
Rigel, absolute magnitude 336; elements *310*; naming 302
Right Ascension (R. A.) 52, 302
Rills, on Moon 129
Ritchey G. W. 331, 399, 401, 407
Ritter 478
Roberts, Isaac 401, 407
Roche 436
Rockets, for observation 76, 90, 92, 473–4
Rocks 107, 111; derived age 428
Rodés, P. Luis 175
Rook Mountains 129
Rösch J. 475
Ross F. E. 189
Ross, James 102
Ross 128, *310*
Ross 154, *319*
Ross 248, *310*
Ross 614, *310*; B, 317
Ross 627, *343*
Ross 780, *310*
Rosse, Lord 395, 401, 455
Rossiter E. 493
Ruduax, Lucien 139, 186, 286
Russell H. N. 358, 432, 440
Rutherford, Ernest 356–7
Rydberg's Constant 482–3
Ryle 475

S

S waves, through Earth 110–1
Sacramento Peak, solar station 475
Sagittae, S, *321*
Sagittarii, Nova, *326*
Sagittarii, Nova I, *326*
Sagittarius 302; nebula 368
Saha M. N. 496
Sandage 445
Saros Cycle 157, 164
Satellites, definition 7; dimensions 63; direction of orbits 62; distribution and number 60; of Jupiter 213; of Mars 205; of Neptune 227; of Saturn 222; of Uranus 226; relative sizes 113
Satellites, artificial, 472–4
Saturn, apparent diameter 218; appearance 220; belts 220; colour 220; constitution 221; density *63*, 216; diameter *63*, 216; distance from Earth 218 from Sun 218; elements *62*; flattening 79, 216; inclination of axis 218; known of old 57; magnitude 304; mass *63*; occultation 168; opposition 218; phase-effect 67; revolution period 218; rotation period 216, 221; seasons 218; shadow of rings 220; spots 220; temperature 221; velocity 218; volume *63*, 216
Saturn's Rings; apparent diameter 218; appearance 220–1; brightness 221; constitution 221; crêpe 221; diameter 218, 221; divisions 221; orientation period 218; rotation period 221–2; size of particles

222; thickness 220, 222; transparency 222
Saturn's Satellites 222–4
Schaeberle 316
Schalen C. 373
Schatzman E. 334
Schaumasse, comet *232*
Schiaparelli 183, 186, 188–9, 200, 204, 250
Schlesinger F. 493
Schmidt B. 456
Schmidt camera 456–7; largest *456*
Schröter 149, 183–4, 188–9, 204
Schwabe 276
Schwarzschild K. 352, 392, 445
Schwassmann-Wachmann I, comet *232*
Schwassmann-Wachmann II, comet *232*
Scoriae, *see* Volcanic activity
Scorpio 302
Scorpii, α, diameter *338*; elements *310*, *313*; galactic longitude 360; spectral type 342
Scorpii, μ, A, *348*; B, *348*
Scorpii, ξ, 314; A, *315*; B, *315*
Scorpii, U, 330
Scorpii, RR, *322*
Scorpii, V 381, 338
Sculptor, nebular system *418*, 419
Sea, surface of equilibrium 74; variation in level 82; *see* Land and Sea
Seares 305, 361, 364
Seas, lunar, *see* Maria
Seasons 282; *see* individual bodies
Secchi, Fr 198, 340, 481
Seismograph 108, 110
Selected areas, of star sphere 304
Serpentis, δ, *313*
Serpentis, RT, 330
Serpentis, Nova, *326*
Seven-tailed comet of 1744, 234
Sèvres, standard metre 79
Shadow, Earth's 86
Shadows, on Moon's surface 124, 128, 145
Shapley H. 323, 383, 385, 399, 400, 412, 415, 444, 491
Shooting stars, *see* Meteors
Short J. 190
Short time-scale, *see* Universe, age
Sial, on Earth 112, 438
Siberian meteorite, of 1908, 252, 255; of 1947, 255
Sidereal day 272
Sidereal revolution period, Moon's 114
Sidereal rotation, *see* Sun
Siderites 256
Siderolites 256
Signs of Zodiac, *see* Zodiac
Silicates, on Mars 200; *see* Sial, Sima
Silicon, in Earth's crust 111, in meteorites 256; in solar atmosphere *351*; in Sun 272
Sima, on Earth 112, 438
Simons, Major David 90
Sinton W. M. 190
Sirius, apparent brightness 304–5; colour 338; companion 316; elements *310*, *315*; first magnitude 336; galactic longitude 360; irregularities in motion 316; luminosity 345; magnitude 304–5; mass 345; naming 20, 302; spectral type 340, 345
Sirius A, *310*, *348*
Sirius B, absolute magnitude *343*; density 343; diameter *338*; 343; Einstein effect in spectrum 494; elements *310*; luminosity 343, 348; mass 343; temperature 343; white dwarf 344
Skjellerup, comet 240
Sky, appearance 3
Slipher V. M. 204, 398, 418
Smith S. 419
Soddy 356
Sodium, in Earth's crust 112, vapour from comet 240
Sodium chloride, in Venus's atmosphere 189
Solar atmosphere, elements *351*; *see* Sun
Solar batteries 284
Solar Constant, *see* Sun
Solar cycle 160; *see* Sun
Solar day 272
Solar flares, *see* Sun
Solar parallax, values *462*
Solar radiation, *see* Sun
Solar rays, *see* Sun
Solar spectrum, *see* Sun
Solar System, composition 60; definition 17; diameter 17; discussion 19–20, 57; distribution of Asteroids 207 of planets 61; essentials 436; extent 60; formation 437; orbital relationships 436–7; orbital velocities 393; origin 146; position in Galaxy 21; scale 17
Solar type stars 340
Solar vortices, *see* Sun
Sommer 296
Somnium, Kepler's 148
Sound, in upper atmosphere 92; speed 17
South Africa, Obs., *455*
Southern Cross, *see* Crux
South Tropical disturbance, *see* Jupiter
Soviet Union, Obs., *455*
Space 85, 290, 292, 297
Space velocity, *see* Stars
Specola Vaticana Obs. 198, *456*
Spectrograph 480
Spectroheliograph 280, 287, 475
Spectrohelioscope 475
Spectroscope 479
Spectroscopic analysis 478, 481; of Earth's atmosphere 160; of eclipsed Moon 160; of Jupiter 212; of Jupiter's satellites 215; of Mars 198, 200; of Moon's atmosphere 137; of Venus 189
Spectroscopic stars, *see* Stars, binary
Spectroscopic classification, *see* Stars
Spectrum, absolute magnitude of stars 311; absorption bands 486–7; analysis, *see* Spectroscopic analysis; atomic 483–5; black body radiator 495; companion of Sirius 499; continuum 482; discovery of solar 478–9; displacement of lines 488; distinction in characteristics 481–2; effect of temperature 496; electrical field 492–4; emission and absorption 481, 483; energy curve 495; flame, arc and spark 482–4; Fraunhofer lines 479–480; gravitational effect 494; identification of elements 481; ionised elements 485; magnetic field 491–2; molecular 485–7; motion of stars 307; night light 296 photographic 481; prismatic and diffraction 479; reversal of lines 481; stellar 481, 484, 487; 490–1; Sun's corona 278; telluric lines in Sun's 489; unknown 485; width of lines 492; *see* Balmer's lines, Colour, Doppler effect, Einstein effect, Forbidden lines, Fraunhofer lines, Helium, Hydrogen, Stark effect, Stars, Sun, Zeeman effect
Spencer Jones H. *462*
Spica *310*
Spiral nebulae, *see* Extragalactic nebulae
Spörer's Law 269
Spurious disc 453
St Andrew's Obs. *456*
St John 189
Stark effect 492, 494
Star catalogues, *see* Stars
Star charts, *see* Stars
Starlight, *see* Night light
Star streams 308, 392–3
Stars, absolute magnitude 336; apparent brightness 304, 335; apparent magnitude 364; appearance 20; as suns 335; atmospheres 349–351; axial rotation 492–3; Baade's characteristics 445; Baade's Populations 435, 444; black dwarfs 447; blue giants 407; brightest *310*; carbon-nitrogen cycle 357–9; Cepheid, *see* Cepheids; charts 302–4; cluster variables 383 reactions 359; colour 304–5, 338; colour indices 377; colour temperature 495–6 of types B, A, F, *496*; composition 349, 350, 352; convective equilibrium 353; degeneration 355; 443–4, 447; densities 343, 349; diameters 196, *338*; discussion 301; distance 17, 306, 308, 347; distribution in Galaxy 361, 363 in extragalactic nebulae 407; double, *see* Stars, binary; dwarfs 338, 342; eclipsing variables, elements *319*; effect of refraction 99; energy 353, 355–6, 428, 439; evolution 427, 432, 439–440, 443–5 of Populations I and II 445; formation 434–5; future 443, 447; galactic concentration *361*; galactic globules, *see* Globules; gas laws 352–3; general nature 8; giants 338, 342; hydrogen into helium 357–9; interiors 350, 352; invisible 316; ionised 352–3; irregularities in motion 316; light variation 319; lives 432; luminosity 335, 344–7; magnetic field 494; magnitudes 304–5; Main Sequence 381, 440, 445, dwarfs 443, extragalactic 411, reactions 357–9; masses 342–3, 348; mass-luminosity relation 348; mean number *306*; measurement of motion 489 of radial velocities 489–491 of temperatures 495–6; minimum visual magnitudes *304*; motions 306–307, 392; movement 10; multiple 312; naked-eye 302–3, 305; near 311 to Sun *310*; novae, *see* Novae; nuclear reactions 357–9; number 303–5 in Galaxy 364 by magnitudes *306*; observation 301; occultation 168; opacity of atmosphere 351; orbits 393; parallax 309, 311, 462; photography 304–5; photosphere 350 as black body 495; planetary systems 317–318; pressure 349; 353; quadruples 314; radial velocities 307, 392–4; radiation 349–350, 352–3; radiative equilibrium 352–353; red dwarfs, definition 338, luminiosity 344, magnitudes 336, reaction 359; red giants 381, 383, formation 445, extragalactic 411, reactions 359; red stars 324; reddening 377, 379; seen from Moon 151; sizes 336–8; solar-type dwarfs 411; spectra of dwarfs 492, 496–7 of giants 492, 496–7 of novae 491, 497 of white dwarfs 494; spectroscopic classification 339–340, 347; standard for comparison 335; super-giants, definition 338, evolution 440, 443–4, extragalactic 407, magnitudes 336, mass and density 343, measurements 461–2; supernovae, *see* Supernovae; systems of naming 20, 302; temperature 305, 338–9, 349, 352–3, 355; triples 314; variation of magnitude with altitude *305*; velocity 306–7; visible number 292; Wolf-Rayet, *see* Wolf-Rayet
Stars, binary: coloured 313; components and magnitude *313*; doubles 311–2; eclipsing binaries 318, spectra 493; mass-luminosity relation *348*; measurements 458–62; optical doubles 312; orbits 314–5 period and eccentricity *315*; resolution 460–3; spectroscopic 313–5, 490, 493; visual 313, 315; wide binaries 432
Stars, variable: semi-regular 324; designation 318; eclipsing, elements *319*; measuring 337; ellipsoidal 319; emission spectra 328; irregular 324–5 elements *326*; long-period 440 elements *322*; nebular 327–8, 330; P Cygni type 329–330; peculiar 327; periodic 320; pulsating 323; semi-regular 324 elements *325*; spectra 491, 497; SS Cygni type 330 period and amplitude *331*; Z Camelopardalis type, period and amplitude *331*
Stars, white dwarf: 338, 432–3, 440, 443, 445, 447; absolute magnitudes *343*; densities 343, 354; dimensions 354; interiors 354–5, 359; luminosity 344; mass 343; structural difference 348
Stebbins J. 377, 386, 411, 477
Stellar, *see* Stars
Stephan-Oterma, comet *232*
Stereo-comparator, *see* Blink microscope
Stevens 90
Stony meteorites 256
Störmer 103, 288
Strand 317
Strassman F. 357
Stratosphere 90, 92, 160; *see* Atmosphere
Strömberg 392; asymmetrical star streams 392–3
Strong J. 190, 455
Stroobant 60
Struve O. 373, 375, 492
Struve 2398, A, *310*; B, *310*
Subatomic energy 356–7
Suess H. 189
Sulphur, in solar atmosphere *351*
Summer lightning 104
Sun, absolute magnitude 335; age 445; angular diameter 19, 260; angular separation from Mercury 181, from Venus 187; antipodal point 66–7; apparent diameter 18; apparent magnitude 335; apparent nature 259; appearance 6; a star 259; atmosphere 264, 272, 275, 277; atmospheric layers 276–7; axis, inclination 269; behaviour relative to Moon 124; Bjerkness's theory of spots 279–280; bodies passing across 60; brightness 262–3; carbon-nitrogen cycle 357–8; celestial latitude 122; centre of solar system 57; chromosphere 154, 262, 265, 272–278, 287; colour 272–3; comparative magnitude 336, 338; composition 111, 272; constitution 263; corona 262, 265, 275, 277–8 at eclipses 167; coronal jets 275; cyclonic theory of spots 278; density 262, 264, 343, diameter 259, 260, *338*, 343; dimensions 259; discussion 259; distance 466 from centre of Galaxy 385; distribution of spots 268–269; effect on Earth's electric field 103, 105 on terrestrial phenomena 281 on Zodiacal Light 294; effects of radiation 499; elements *310*, *348*; energy output 355, 440; evolution *440*; faculae 266–8, 273–4; filaments 275; flares 89, 262, 275, 287–8; flattening 260; flocculi 274–5; formation 445; future 446–7; helium radiation 358; hydrogen into helium 357–8; illuminating Moon 116, 128; planets 66; in region of star density 390; limb darkening 351; magnetic field 491, 493–4; Main Sequence dwarf 350, 353, 355; mass 262, 343; nuclear reaction 352, 357; observation 475–6; orbit 393; parallax 309, 461–2, 489; period of violent activity 160; periodic variations of atmosphere 273; photosphere 262–3, 265, 279, 350–1; pores 267; position in Galaxy 363, 385–6, 393, 415–6; pressure 350; probable evolution 440; prominences 167; 273–4, 277; proper motion 307–8; radiation 167, 262, 350, 356 and corpuscular emissions 281, 286, 288, of corona 278; warming Earth 109; radio observation 262; rays 6; real motion 393; reflected by Moon 124; relationship to Galaxy 21; relative abundances *351*; relative positions to Moon 121; reversing layer 265, 272, 481, 496; rice-grain 265; rotation 269, 272, 280, 288 in Galaxy 394; rotation period 272, 489 speed 492; seen from Moon 154; solar constant 283, 288; solar cycle 273, 275–6, 283, 288; solar vortices 275; spectrum 97, 481, 496 of corona 278; spectral type 342; spots 262, 266–269, 273, 275–6, 278–280, 286, 491, 493, 498; spot groups 268; sunspot cycle, *see above* solar cycle; surface gravity 262, 350, 494; temperature 263–4, 350–3 of corona 278 of spots 267; tidal effect 170; transits of Mercury 182 of Venus 187, 189; ultra-violet light 189; radiation 288; use of solar heat 283–4; using eclipse 60; variations in diameter 260; velocity 308; volume 259, 262
Sun-dogs, *see* Parahelia
Sunrise and sunset 73–4, 100–1
Sunspots, *see* Sun
Sunspot stars 342
Supernovae, description 331; discovery 332; elements of extragalactic *333*; formation 444; frequency 333; in Crab nebula 334; in IC 4182, 334; in NGC 4321, 333; in Ophiuchus, 334; light curves 333; magnitudes 333; origins 334; spectra 333; Tycho's star 333; types 333; *see* Stars
Super planets 317
Surface, of Sun and stars 349
Surface of compensation 112
Surface of equilibrium, of sea 74
Sverdrup, Capt. O. 124
Synodic period, definition 71; Moon's 115; *see* Sun, rotation
Syzygy 170, 172

T

Tail, of comet 234, 236–7, 240
Tangential direction 16
Tauri, α, diameter *338*; elements *310*; magnitude 323; spectral type 342; *see* Aldebaran
Tauri, λ *319*
Tauri, RV, 324
Taurus, moving clusters 308
Telescopes, astronomical 451;

drive 457; electronic 469—470; focal lengths 459; invention 57, 451; limitation 177, 193; mounting 457; principle of reflecting 454—5, of refracting 452, 453; radio-telescopes 92, 471—2; resolving power 453, *454*; Schmidt 456; size of refractors 453; solar 471—2; use of refractors 458; world's largest reflectors *455* refractors *454* Schmidt *457*
Teller 435, 440
Telluric currents 105
Telluric lines, *see* Spectrum
Tempel I, comet *232*
Tempel II, comet *232*
Tempel's Comet 250
Tempel, Swift, comet *232*
Temperature, properties 90—91; *see* Earth, Moon, planets, Stars, Sun
Temporary stars, *see* Novae
Terminator, on Moon 124, 126, 129; *see* planets
Tethys *222*
Theophilus, crater 133
Thermocouples 471—2
Thibaud J. 356
Thiessen G. 494
Thomson J. J. 356
Thorium, on Earth 105, 110
Thunder, cause 104
Tidal effects, of Moon 118; of Sun 170
Tides, action 174; amphidromic 172; amplitude 170, 174; bore 174; bulge 170, 172; crustal 172; effect of quadratures 172 of syzygies 172; energy 174; erosion 174; establishment of the port 172; geographical differences 172; highest and lowest 172; in relation to Earth's core 106; interval 172; mechanism 170; prediction 172; range 172; spring and neap 170, 172; theory 169—170
Tikhoff G. A. 191
Time-scales, long and short 430, 432, 439
Titan *222*, 223—4
Titania *225*
Titanium, on Earth 112
Titanium oxide, stars 342; vapour in star atmospheres 324
Titius 62
Tolman 424, 426
Tombaugh, Clyde 59
Tonanzintla Obs. *457*
Transit, *see* Mercury, Venus
Trianguli, R. *322*
Triton *227*
Tropopause 90; *see* Atmosphere
Troposphere 90—1; *see* Atmosphere
Trumpler R. J. 381; coefficient 381
Tucanae, 47, 382—3
Tuttle, comet *232*
Tuttle-Giacobini, comet *232*
Tuttle's Comet 250
Twilight 86, 90, 137
Tycho's Star 333

U

Uccle Royal Obs. 59, *456*
Ultra-violet photographs, of Mars 195, 197; of Venus 189
Ultra-violet radiation, decomposing influence 90; discovery 478; from Sun 89, 240, 286, 288; in galactic nebulae 369—370; on to Moon 148, 154
Umbra, definition 158; Moon passing through 158—9; of sunspots 266
Umbriel *225*
Universe, age 427, 430, 440; constitution 424; evolution *435*; expansion 422, 424, 426, 428, 430, 432—4, 475; future 446; limits of observation 426; mass 433; origin 427, 433—4; pre stellar 433—4; primal atom 433—4; radius 424, 434; time-scale 430; visible 21; *see* Metagalaxy
Uranium, disintegration *428*; in Earth's crust 105
Uranus, apparent diameter 224; appearance 224; atmosphere 226; belts 224; colour 224; constitution 226; density *63*, 224; diameter *63*, 224; discovery 57, 59, 62; distance from Sun 224; eccentricity of orbit 224; elements *62*; flattening 224; inclination of axis 224 of orbit 224; magnitude 304; mass *63*; opposition 224; perturbation of orbit 57; revolution period 224; rotation period 224; seasons 224; temperature 226; velocity 224
Uranus's satellites 60, *225*, 226
Ursa Major, clusters of galaxies *419*; moving cluster 432
Ursae Majoris, ζ, *313*; ζ-, *315*; *see* Mizar
Ursae Majoris, φ, *315*
Ursae Majoris, W, 319; elements *315*, *319*; luminosity,, mass, spectral type 346; A, *338*
Ursae Minoris, α, elements *321*; spectral type 340

V

Väisälä, comet *232*
Van Albada 433
Van Allen belt, 472
Van de Kamp 317
Van der Hulst 473
Van Maanen's Star 306, *310*
Van Maanen 2, *338*, *343*
Van Rhijn 391
Variable stars, *see* Stars
Vassy, M and Mme 296
Vatican Obs., *see* Specola Vaticana
Vega, colour 338; elements *310*; naming 20, 302; spectral type 340
Vegetation, *see* Mars
Velocities 12; *see* Earth, Moon, planets, Stars, Sun
Velorum, γ, 340
Venus, albedo 187; angular diameter 463; angular separation 187; apparent diameter 18, 187; apparent movement 66, 69; appearance 187; Ashen Light 190; atmosphere 187—9; brightness 187; charts 188; clouds 187, 189; colour 191; conjunction 187; density *63*, 186; diameter *63*, 186; distance from Earth 187 from Sun 186; eccentricity of orbit 186; elements *62*; elongation of orbit 187; inclination of axis 189 of orbit 187; inferior planet 66; infra-red examination 189; known of old 57; life 189, 191; markings 187, 189; mass *63*, 186; observation 57, 60, 66—67; occultation 168; orbit 186; orbital velocity 393; parallax 463; phases 66, 187; polar compression 186; polarisation of light 189; radio waves 189; resemblance to Earth 186; rotation period 188—9; satellite 190; sidereal period 188; spectroscopic examination 189; surface 187—8, 190; temperature 189—190; transits 187, 189; ultra-violet examination 189; volume *63*, 186; water-vapour 189—190
Vertical, Earth's 72, 107
Vesta 63, 205, 208
Vienna Obs. *454*
Vigroux 160
Villard P. 103
Virginis, α, *310*; *see* Spica
Virginis, γ, *313*, *315*
Virginis, 78, 494
Virgo, cluster of galaxies 412, 419; condensation 434; No. 1, *419*
Visibility 73, 76—8; in darkness 291; refraction 99
Vogel 489
Volcanic activity 89—90, 107—8, 161
Volume of planets *63*
Vulcan, supposed observation and position 60
Vulcanism, *see* Volcanic activity

W

Waldmeier 276
Waldemeier 476
Wandering stars 57
Warner and Swasey Obs. *456*
Washington Catalogue 5583 and 5584, 307
Washington Obs. 205, *454*
Water, affecting electric field 104; influence on Earth's surface 148; on Mars's surface 198; on Moon 128—9, 136, 146, 148; on Venus's surface 191; unique proportion on Earth 82; vapour condensing 92, 94; vapour diffusing light 97; vapour forming clouds 92 in air 88 in atmosphere 88, 97; vapour in atmosphere of Mars 198 of Venus 189—190 in troposphere 90; *see* Land and sea, Tides
Wave mechanics 483
Weather 91; *see* Clouds, Lightning
Weight 15, 81—2
Weizsäcker C. von 357, 433, 437, 439
Westphal, comet *232*
Whipple F. L. 189—191
Whipple, comet *232*
White dwarfs, *see* Stars, white dwarf
White nebulae, *see* Extra-galactic nebulae
Whitford 386, 411, 469, 477
Wiechert 111
Wildt R. 189, 351
Wilson O. C. 343
Wirtanen, comet *232*
Wolf, Max 62, 206, 371—3
Wolf I, comet *232*
Wolf 219, *343*
Wolf 359, *310*, *348*
Wolf 424, A, *310*; B, *310*
Wolf 457, *338*, *343*
Wolf 489, *343*, *344*
Wolf 1346, *343*
Wolf-Rayet stars, formation 444, spectra 497; type 340
Wolf's sunspot number 262
Wollaston 479
Wood 479
Wright, Thomas 361
Wylie C. C. 256
Wyse 416

X

Xenon, in air 88
Xenophanes 148

Y

Yale University Southern Station, *see* Johannesburg
Yamamoto's Catalogue 233
Yerkes Obs. 59, 329, 347, 373, 401, 453, *454*
Yntéma 292
Young 460, 479, 489
Young's wave theory of light 479

Z

Zeeman 491
Zeeman effect 491—4
Zirconium oxide stars 342
Zodiac, Signs of 302
Zodiacal belt, Moon's orbit 168
Zodiacal Light 292—4, 296
Zöllner 464
Zone of avoidance 377, 419
Zones of Earth, climatic 42
Zwicky F. 333—4, 419, 421, 432, 444, 477

ASTRONOMY